

Intuition, deduction, and the art of picking up the pieces

Citation for published version (APA):

Senden, M. (2016). *Intuition, deduction, and the art of picking up the pieces: reciprocal gains of computational modeling and fMRI*. [Doctoral Thesis, Maastricht University]. Ridderprint. <https://doi.org/10.26481/dis.20161104ms>

Document status and date:

Published: 01/01/2016

DOI:

[10.26481/dis.20161104ms](https://doi.org/10.26481/dis.20161104ms)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

**Intuition, Deduction, and the Art of Picking Up the Pieces:
Reciprocal Gains of Computational Modeling and fMRI**

Mario Senden

Intuition, Deduction, and the Art of Picking Up the Pieces: Reciprocal Gains of Computational Modeling and fMRI

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public on
Friday, 4th November 2016, at 10.00 hours

by

Mario Senden

Supervisor

Prof. dr. R. Goebel

Co-supervisor

Prof. dr. G. Deco (Universitat Pompeu Fabra, Barcelona, Spain)

Dr. J. Peters

Assessment Committee

Prof. dr. P. De Weerd (chair)

Prof. dr. P. Roelfsema (Netherlands Institute for Neuroscience, Amsterdam, The Netherlands)

Priv. Doz. dr. P. Ritter (Charité, Berlin, Germany)

Dr. A. H. van der Lugt

Dr. V. G. van de Ven

CONTENTS

Chapter 1	General Introduction	7
Part I	Theory Testing	19
Chapter 2	Evaluating Population Receptive Field Estimation Frameworks in Terms of Robustness and Reproducibility	21
Chapter 3	Comparing Computational Models of Simultaneous Brightness Contrast Using Ultra-High Field Functional Magnetic Resonance Imaging	49
Part II	Theory Building	75
Chapter 4	Rich Club Organization Supports a Diverse Set of Functional Network Configurations	77
Chapter 5	The Rich Club Can Organize State-Dependent Functional Network Formation by Engaging in Oscillatory Behavior	95
Chapter 6	General Discussion	123
	References	133
	Appendices	147
	Valorization	169
	Acknowledgements	173
	Publications	175
	Invited Talks	177
	Curriculum Vitae	179

General Introduction

Chapter

1

The most humble experience is already, in itself,
structured like a thought.

Ibn Tufail

PRELUDE

In the 12th century allegorical novel “Ḥayy ibn Yaqẓān” (English: “The Improvement of Human Reason: Exhibited in the Life of Hai Ebn Yokdhan”) the Andalusian philosopher Ibn Tufail tells the story of feral child Ḥayy. Despite having lived on a deserted island and being raised by a gazelle, Ḥayy grew up to be a learned man. What initially set him on his path towards knowledge was the death of his surrogate mother and his resolution to identify and remove whatever caused her to assume that state. Observing that outwardly she appeared to be unharmed, he concluded that her injury had to be internal. He further realized that the cessation of function was not limited to a single part of her body but affected it in its entirety and so concluded that the part he was looking for was integral to the functioning of her body. He decided to examine her inner organs concentrating on those found within the chest. Doing so, Ḥayy eventually arrived at the heart and believed it to be what he was looking for. Upon examining the heart he found that it had two cavities and that the left was empty while the right contained congealed blood. Conjecturing that the left cavity could not be without purpose, he came to the conclusion that it must be the seat of the life force and that his mother’s life force had subsided. After realizing that while his mother’s body appeared to be intact, her essence was gone, Ḥayy stopped his inquiry into his mother’s death. Instead, he began to examine the island around him employing the same approach of systematic observation and experimentation. He monitored the behavior of animals, engaged in thought experiments to inquire into the nature of the heavenly bodies, and proposed general laws which govern and explain all specific events around him. Eventually, after a lifetime of reasoned inquiry, he arrived at an understanding of the world and of himself.

The story of Ḥayy exemplifies the workings of a scientifically inclined mind and stresses the importance of reason and observation as complementary tools to understand nature. It is a story immediately appealing to any scientist as it illustrates two widely accepted characteristics of the scientific method (Whewell, 1858): 1) a theoretical construct which allows us to make sense of what we perceive and 2) systematic observation which allows us to test our theories. Neither of the two aspects suffices individually, of course. It is their interplay which is essential for any scientific endeavor to be successful. This can be recognized by tearing the two apart. A theoretical construct which is not constantly evaluated against reality through systematic observation leads us astray no matter how elegant or internally consistent it might be. Such a situation is typical for pseudoscience which is often characterized by large, albeit not necessarily consistent, theoretical constructs in conjunction with either reluctance to or build-in immunity against empirical testing. On the other hand, empirical observations which are not organized by theory are nothing but a conglomeration of unintelligible data. Some organization of this data might still be achieved through data-driven analyses which render the data heap practically useful. Genuine understanding, however, remains elusive unless

comprehensible theories are developed which can not only predict but also explain relations among observations and phenomena.

The importance of the interplay of theory and observation has, of course, been recognized by many philosophers over the centuries and has been most beautifully expressed in Kant's epistemological synthesis "thoughts without intuitions are empty; intuitions without thoughts are blind" (Kant, 1787). Hence, progressive scientific disciplines always entail an element of theory and observation (Lakatos, 1976) even if the emphasis is placed more on one or the other, depending on the discipline. However, the development and increased utilization of high throughput methods used to accumulate "big data" might threaten this essential interplay of theory and observation (Krohs, 2012; Leonelli, 2012). While "big data" science surely has many advantages, it carries with it the temptation to effortlessly produce large amounts of publishable data and to neglect the accompanying development of comprehensible theories with the hope that the data will "speak for themselves" (Allen, 2001; Krohs, 2012; Leonelli, 2012). Within the neurosciences, the field of neuroimaging has due to its most popular method, functional magnetic resonance imaging (fMRI), become a "big data" science (Van Horn & Toga, 2014). Recognition of the symptoms of this status has led some researchers to argue that functional neuroimaging had so far not informed theory in any meaningful way and to suggest that it might never will (Coltheart, 2006; Page, 2006).

This is too harsh and too quick a verdict, however. While fMRI might on occasion be misused for "convenience experimentation" (Krohs, 2012), there is no inherent flaw in the method preventing it from being, and indeed having been, used in a theory-driven way (Chatham & Badre, 2014). Nor is there a lack of interest within other disciplines such as psychology (Chatham & Badre, 2014) or computational neuroscience (Peters, Jans, van de Ven, De Weerd, & Goebel, 2010; Peters, Reithler, & Goebel, 2012; Rolls, Grabenhorst, & Deco, 2010; Tagamets & Horwitz, 1998) to build and test theories in the context of fMRI research. A liaison between neuroimaging and computational neuroscience appears to be especially promising given the latter's richness in theory and focus on neural plausibility (Boone & Piccinini, 2015; Rescorla, 2015). While this liaison is promising and clearly desired (Corchs & Deco, 2002; Peters et al., 2012; Rolls et al., 2010; Tagamets & Horwitz, 1998), interdisciplinary work always presents with a number of challenges (Jeffrey, 2003). The challenges for the integration of fMRI research and computational neuroscience stem from differences in their focus with respect to a) spatial and temporal scales and b) levels of description of information processing systems according to Marr's Tri-Level hypothesis (Marr, 1982). Thus in order to understand the challenges as well as opportunities for the integration of the two disciplines it is necessary to analyze each in terms of these two factors.

SPATIAL AND TEMPORAL SCALES

A first insight into the aforementioned challenges comes from an analysis of fMRI research, computational modeling, and neuroscience in general in terms of spatial and temporal scales. Spatially, neuroscientific research can be conducted at what Churchland and Sejnowski (1992) termed the brain's "levels of organization". These include the microscopic, the mesoscopic, and the macroscopic level. At the microscopic level, which is in the range of about $100\mu\text{m}$, research concerns itself with small clusters of interconnected neurons together forming cortical microcircuits (Lorente de Nó, 1949; Silberberg, Grillner, LeBeau, Maex, & Markram, 2005). At the mesoscopic level, which is in the range of millimeters to a few centimeters, the main focus of research is on cortical columns and laminae. Cortical columns are characterized by larger groups of vertically arranged neurons with similar response profiles (Hubel & Wiesel, 1959; Mountcastle, 1997; Szentágothai, 1983), whereas cortical laminae are horizontal organizations of neurons of similar cell type and diverse connectivity profiles (Brodmann, 1909). Finally, research at the macroscopic level investigates brain function at scales ranging from local maps over regions to whole-brain dynamics.

Temporally, the scales at which neuroscientific research is carried out depend on the temporal resolution of different methods but also on temporal aspects of brain activity such as rhythmicity. With regard to the latter, scales are dictated by 1) spike timing and intra-cellular rhythms in the range of a few milliseconds, 2) synchronization of local cell assemblies in the range of up to 100 milliseconds, 3) large-scale information transfer and integration in the range of several hundred milliseconds (Le Van Quyen, 2011), and 4) infraslow fluctuations in the range of several seconds (Vanhatalo et al., 2004).

When applying this analysis to computational neuroscience and fMRI it becomes quickly apparent that they largely operate at different scales. Computational neuroscience arguably operates for the most part at the micro- and mesoscopic spatial scales whereas fMRI largely operates at the macroscopic scale. This does not imply that computational neuroscience does not attempt to explain behaviorally relevant processes or does not propose models of entire cortical regions or pathways. Rather, the most informative testable predictions these models make with regard to information processing in the brain can best be evaluated at micro- and mesoscopic scales. With regard to the temporal dimension a similar chasm exists between the two fields. Computational neuroscience tends to produce simulated data with high temporal resolution whereas fMRI is limited by the sluggishness of the blood oxygen level dependent (BOLD) signal (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001) and thus often does not offer the temporal detail needed to evaluate computational models. This is not to say that generating predictions of brain activity as revealed by fMRI could not or were not generated from computational models (e.g. Rolls et al., 2010). Nevertheless, data from electrophysiological methods which operate at smaller spatial scales and offer higher temporal resolution were generally better suited to inform and evaluate computational models.

With regard to the spatial dimension this state of affairs is beginning to change as a consequence of recent technological advancements. With increasing power of computers in general combined with the increased use of massively parallel computing employing graphic processing units (GPUs; Nageswaran, Dutt, Krichmar, Nicolau, & Veidenbaum, 2009; Woodman & Jirsa, 2013) and the advent of dynamic mean field models of neuronal populations (Wong & Wang, 2006) and other neural mass models (Freyer et al., 2011; Freyer, Roberts, Ritter, & Breakspear, 2012), computational neuroscience can now build models of ever increasing brain coverage without sacrificing necessary spatial detail. At the same time, developments relating to ultra-high field fMRI push the limits of spatial resolution deep into the mesoscopic scale (Duyn, 2012). Ultra-high field fMRI now permits spatial resolutions sufficiently high to study brain function at the level of cortical columns (e.g. Yacoub, Harel, & Uğurbil, 2008; Yacoub, Shmuel, Logothetis, & Uğurbil, 2007; Zimmermann et al., 2011) and laminae (e.g. Muckli et al., 2015). In terms of temporal resolution, current developments have not yet led to major improvements for non-invasive human fMRI research. However, by utilizing increased computational power, large-scale models exhibiting temporal dynamics in the range observable by fMRI can bridge the gap from the other side.

These developments are significant and exciting as they open up the possibility of integration of computational neuroscience and fMRI at two scales. First, ultra-high field fMRI allows for testing detailed spatial model predictions at the mesoscopic scale. Second, increased computational power as well as recently developed dynamic mean field neural population models allow for the development of models relating to global brain dynamics where their predictions can be tested by standard spatial and temporal resolutions of fMRI.

TRI-LEVEL HYPOTHESIS

A second informative analysis of neuroscience is in terms of levels of description of information processing systems commonly referred to as Marr's Tri-Level hypothesis (Marr, 1982). David Marr and Tomaso Poggio were among the first to realize that an information processing system such as the brain needs to be understood at several distinct yet complementary levels (Marr & Poggio, 1976; Marr, 1977). These authors observed that the most prominent fields during the 1960s and '70s, neurophysiology and psychophysics, though providing detailed and important descriptions of the behavior of neurons and subjects, respectively, generally did not explain either what the purpose of such behavior was or how it was brought about. Marr, in his book *Vision* (Marr, 1982), subsequently provided an analysis of neuroscience which would allow researchers to go beyond description of phenomena. Specifically, Marr proposed three levels at which an information processing system must be understood: at the level of the computational theory, at the level of representation and algorithm, and at the level of physical realization.

Computational Theory. The level of computational theory is the most abstract of the three levels and is concerned with the question of the purpose of any given computation. Understanding this purpose would then greatly aid in identifying and understanding the specific mechanism by which this goal is realized. At this level it is not sufficient to merely state the final goal of the computation, however. It is equally important to understand why the brain would pursue this goal and by which general logic or strategy it goes about this task. An example of an analysis at this level is Chomsky's (1964) theory of transformational grammar. Chomsky realized that the purpose of grammar must be to provide a system of finite rules which is nonetheless able to interpret infinitely many sentences. The reason *why* he regarded grammar to serve this specific purpose was the ability of humans to utter and understand infinitely many sentences while any information processor (and hence the brain) can only operate using a finite set of rules. Finally, Chomsky gave an account of the general logic of his theory in terms of transformations between deep and surface syntactic structures by the application of which any sentence could be decomposed and understood. Importantly, the theory gives a meaningful account of grammar without specifying the detailed mechanisms by which the proposed transformations are achieved. Indeed, an analysis in terms of these mechanisms would take place at the second level described by Marr.

Representation & Algorithm. This level concerns itself with the question of how a computational theory can be implemented in terms of specific representations of information and algorithms manipulating this information. A representation is a formal system of information encoding whereas an algorithm describes a process by which one representation (input) is mapped onto another (output). A straightforward example of a representation is given by numeral systems. A certain quantity such as seven, for example, can be represented in different ways such as the Arabic or Roman numeral systems. Both "7" and "VII" represent the same quantity but in different ways and the choice of representation has important consequences for the manipulation of that information. Multiplication, for example, is comparatively easy using the Arabic system but far less so when using the Roman system. Such considerations are also important in neuroscience where it is important to know which combination of features might be used to code, for instance, for objects in the visual domain. The way we assume information to be stored in the brain and mental processes to be represented necessarily affects the kinds of algorithms suited for the manipulation of that information.

Physical Realization. The final and most concrete level concerns itself with the question of how a given representation and algorithm can be realized in a physical or biological system. At this level hardware constraints need to be taken into account and it is important to have information regarding the elementary computational principles of the underlying architecture. This includes knowledge of whether the system processes information serially or in parallel.

When considering disciplines within neuroscience, neuroanatomy is clearly linked to the third level. Electrophysiological research is also partly related to this level. Such as when investigating the mechanisms underlying action potentials. But electrophysiology lies at the junction of the third and the second level as it can also provide information on the kinds of representations used by neurons and small neuronal circuits. Psychophysics, on the other hand, is immediately linked to the second level. Every combination of a specific representation and algorithm solves a problem in a certain way and will fail and produce erroneous or illusory results if pushed to the limit. Psychophysical experiments are aimed at pushing these limits and can thus suggest as well as rule out representations and algorithms.

With regard to the fields most relevant to the present thesis, computational neuroscience can be related to all three levels though it mainly operates at the second and third. It is strongly related to the second level as it concretizes computational theories relating to cognitive function by suggesting the specific representations and algorithms which can bring about this function. It is also strongly related to the third level as it emphasizes biological plausibility and hence puts constraints on the hardware implementation of representations and algorithms. Functional magnetic resonance imaging is classically used to inform the first and the third level. At the first level its localization of function provides information on the general logic by which a cortical module achieves a certain goal. For example, the goal of a neural reward system likely is to keep track of the expected value of a stimulus in order to meaningfully guide behavior. The finding that reward magnitude and reward probability are localized in different parts of the ventral striatum (Yacubian et al., 2007) supports this notion. It also suggests that the general logic employed by the brain to keep track of expected value is to keep track of the elements from which it can easily be computed. Data provided by fMRI is thus informative with regard to the computational theories of reward processing. At the third level, fMRI is informative at a macroscopic scale due to its large field of view which provides an insight into the interaction of several cortical regions when performing a task. This makes it possible to distinguish local from global implementations of cognitive processes. Some processes might, for instance, be implemented mainly within a single cortical region while others are the result of a concerted interplay among many regions. Importantly, fMRI can reveal that what might at the first level be described as a single cognitive module may still be neurologically implemented as a globally distributed process. With the advent of ultra-high field fMRI and clever analyses techniques such as multi-voxel pattern analysis (Pereira, Mitchell, & Botvinick, 2009) or representational dissimilarity analysis (Kriegeskorte, Mur, & Bandettini, 2008), fMRI is now additionally pushing for the second level of Marr's description. This is due to the fact that these techniques make it possible to unravel the specific representations of a brain region. While regular univariate fMRI might, for instance, reveal that face processing is localized in the fusiform gyrus (e.g. Kanwisher, Stanley, & Harris, 1999), multivariate and or ultra-high field fMRI can reveal the specific features coded in such a region and can hence be used to inform and evaluate computational models. Additionally, ultra-high field fMRI may also be informative

with respect to the detailed spatial arrangement of representations thus providing constraints on the implementation (3rd level) of computational models at the mesoscopic scale.

This analysis shows that whole-brain models in computational neuroscience can be informed by standard univariate fMRI which provides insights regarding the global computational theory as well as on large-scale biological constraints for model implementation. Yet, at the level of representation and algorithm as well as of implementation at the mesoscopic scale, which are highly important for the evaluation of computational models, standard univariate fMRI is of little help. Fortunately, this gap is bridged as the scale gap narrows and with the development of new analysis techniques.

PURPOSE AND OUTLINE OF THIS THESIS

The preceding analyses have shown that conducting theory-driven research by integrating computational neuroscience with fMRI is challenged by differences with respect to the typical spatio-temporal scales and levels of description each discipline is associated with. However, they have also shown that, due to advances in both fields, no insurmountable obstacles exist to render this endeavor impossible. Ultra-high field fMRI and multivariate analysis techniques are increasingly informative with regard to the second and third level of description at mesoscopic spatial scales and hence move fMRI closer to the realm of computational neuroscience. At the same time, large-scale (macroscopic) models suitable to probe the first and third levels of description move computational neuroscience closer to the realm of fMRI. Drawing on these developments, the purpose of the present thesis is to illustrate two approaches to conduct interdisciplinary research and how each discipline may benefit as a result. Since theory-driven research need not be exclusively aimed at theory testing but can be employed for theory building as well, one approach is presented within the context of the former while the second approach is presented within the context of the latter. The thesis is thus split in two parts with each detailing research dedicated to one combination of approach and aim.

The first part presents research embedded within the neuroscience of vision and relies on advances with regard to ultra-high field fMRI and computational neuroimaging (Wandell & Winawer, 2015) in order to perform comparative theory testing. Specifically, the first part is dedicated to demonstrating that fMRI can be used to compare competing computational models formulated at the second (representation & algorithm) of Marr's levels with respect to their implications for a detailed spatial activation profile. To establish correspondence between computational units in retinotopically organized models of early visual cortex and brain voxels, population receptive fields (pRFs; Dumoulin & Wandell, 2008) are utilized as a bridging principle. An important assumption of this approach is that pRF properties remain constant across time and stimulation procedure. Additionally, pRFs need to be estimated with a high degree of precision. Therefore, chapter 2 evaluates the robustness and reproducibility

of pRF estimates and identifies an estimation framework guaranteeing a good degree of precision in mapping. Subsequently, chapter 3 details how using pRF mapping, empirical and simulated data can be represented within the same reference space, termed the common brain space (Goebel & De Weerd, 2009; Peters et al., 2010, 2012). Chapter 3 furthermore shows how the common brain space allows for the comparison of different computational models exemplified for the phenomenon of simultaneous brightness contrast (Chevreul, 1839; Hess & Pretori, 1894).

The second part presents research aimed at understanding structure-function relationships in the brain. It exploits advances in parallel computing and neural mass models to build large-scale models which can be informed by fMRI data. In accordance with Marr's computational level, the first question addressed in this part concerns the functional role of the cortical rich club, a set of highly connected brain regions which are also densely connected among themselves (Colizza, Flammini, Serrano, & Vespignani, 2006; van den Heuvel & Sporns, 2011; Zamora-López, Zhou, & Kurths, 2009). Chapter 4 utilizes the massively parallel computing capabilities of GPUs to semi-analytically study the implications of different network types in a steady-state attractor model. The results suggest that the rich club may be relevant for the integration of information processing carried out by specialized brain regions as it affects the size and shape of a system's functional repertoire, such as the brain's functional network configurations. Chapter 5 follows up on these results and investigates the general strategy through which the rich club might organize functional network formation. Here, cortical regions are represented by a neural mass model able to exhibit dynamics ranging from asynchronous fluctuations to oscillations. Using this model, it is shown that oscillations constitute a mechanism sufficient to bring about structured network behavior. Furthermore, by adjusting local oscillatory behavior in order for the model to reproduce fMRI data obtained during a range of functional states, it is shown that the mechanism is differentially employed by rich club regions depending on the functional state. Thus, combining large-scale computational modeling with fMRI measurements enabled the research presented in chapter 5 to go beyond presenting a mechanism suitable to shape functional network formation and to additionally provide evidence that the brain's rich club regions do in fact employ this mechanism.

Theory Testing

Part



Chapter

2

Evaluating Population Receptive Field Estimation Frameworks in Terms of Robustness and Reproducibility

This chapter is an adapted version of the following article:

Senden, M., Reithler, J., Gijzen, S., & Goebel, R. (2014). Evaluating Population Receptive Field Estimation Frameworks in Terms of Robustness and Reproducibility. *PLoS ONE*, 9(12)

ABSTRACT

Within vision research retinotopic mapping and the more general receptive field estimation approach constitute not only an active field of research in itself but also underlie a plethora of interesting applications. This necessitates not only good estimation of population receptive fields (pRFs) but also that these receptive fields are consistent across time rather than dynamically changing. It is therefore of interest to maximize the accuracy with which population receptive fields can be estimated in a functional magnetic resonance imaging (fMRI) setting. This, in turn, requires an adequate estimation framework providing the data for population receptive field mapping. More specifically, adequate decisions with regard to stimulus choice and mode of presentation need to be made. Additionally, it needs to be evaluated whether the stimulation protocol should entail mean luminance periods and whether it is advantageous to average the blood oxygenation level dependent (BOLD) signal across stimulus cycles. By systematically studying the effects of these decisions on pRF estimates in an empirical as well as simulation setting we come to the conclusion that a bar stimulus presented at random positions and interspersed with mean luminance periods is generally most favorable. Finally, using this optimal estimation framework we furthermore tested the assumption of temporal consistency of population receptive fields. We show that the estimation of pRFs from two temporally separated sessions leads to highly similar pRF parameters.

INTRODUCTION

An essential aspect of vision research using functional magnetic resonance imaging (fMRI) is the investigation of retinotopic organization of visual cortex (Dumoulin & Wandell, 2008; Lee, Papanikolaou, Logothetis, Smirnakis, & Keliris, 2013). Phase encoded retinotopic mapping as pioneered by Sereno *et al.* (1995) already allowed for the systematic investigation of polar angle and eccentricity properties of visual cortex. Recently, the advent of the population receptive field (pRF) mapping approach, first described by Dumoulin and Wandell (2008), has supplemented knowledge of receptive field location with insight regarding their size and shape. Beyond the immediate scientific interest in receptive field properties, knowledge of receptive fields is crucial for a number of applications. Receptive fields a) provide a source of information for the reconstruction of stimuli from the blood oxygenation level dependent (BOLD) signal (Thirion *et al.*, 2006), b) can serve as target for transcranial magnetic stimulation (Sack *et al.*, 2009), c) assist function based alignment, d) provide a spatial forward model for computational models (Peters *et al.*, 2012), and e) might give important insights with respect to theories of attention (E. DeYoe, Puckett, & Ma, 2013) as well as into pathologies of the visual cortex (Binda, Thomas, Boynton, & Fine, 2013; Dilks, Baker, Peli, & Kanwisher, 2009; Haak *et al.*, 2014) and brain development.

For the utilization of receptive fields for any of the aforementioned applications, it is necessary that they are measured, or rather estimated, with a high degree of precision. For a number of applications it is necessary to estimate pRF parameters from one set of stimuli and use

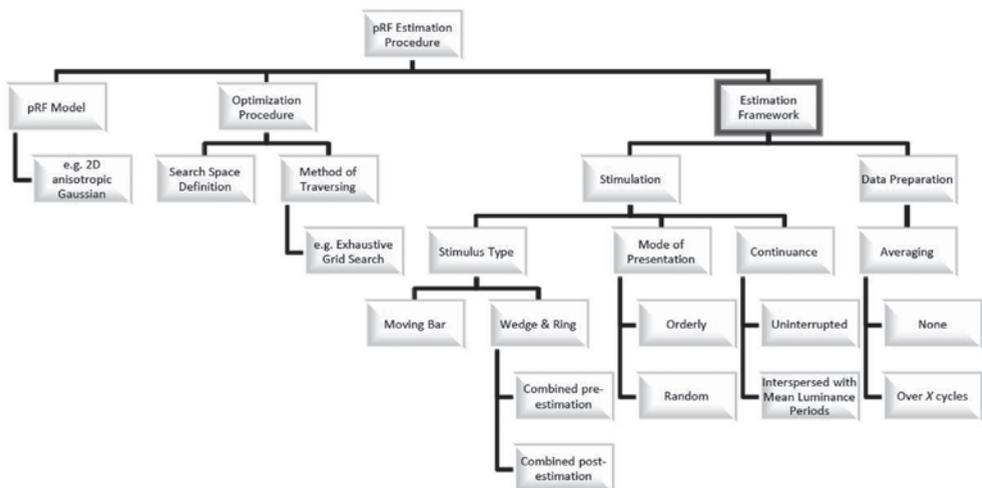


Figure 2.1: Conceptual Division of the pRF Estimation Procedure. This figure gives a visual organization of the pRF Estimation Procedure with its components: 1) pRF Mode, 2) Optimization Procedure, and 3) Estimation Framework. The frame around Estimation Framework indicates that choices related to this component and its subcomponents constitute the focus of this paper.

their predictions on a distinct set of stimuli. For other applications it is necessary to perform estimation in one session to be able to use the obtained parameters in future sessions. For these reasons, the precision of pRF estimation necessarily pertains to generalizability across stimuli as well as across sessions (i.e. time).

Such precision relies first and foremost on three aspects of the estimation procedure (see figure 2.1 for a visual representation of this organization). First of all, an adequate model description of a receptive field is needed to capture its position, shape, and local properties. Secondly, it is necessary to define an accurate and fast optimization procedure by specifying both a suitable search space on the parameters and an efficient way to traverse this space. Finally, it is necessary to set up an adequate estimation framework providing the data from which receptive fields are estimated. With regard to model description and optimization procedure other groups have previously produced excellent work (Dumoulin & Wandell, 2008; Kay, Winawer, Mezer, & Wandell, 2013; Lee et al., 2013). According to our knowledge, however, the estimation framework has so far not undergone a thorough investigation and it is our aim to provide a first attempt by studying how different choices affect pRF estimation performance.

In detail, the estimation framework encompasses stimulation and data preparation. Stimulation, in turn, refers to the type of stimulus used (e.g. moving bar), stimulus presentation (the stimulus changes location in an orderly or random fashion), and continuance (stimulation is continuous or interspersed with mean luminance periods). Importantly, the stimulus type does not necessarily refer to a single stimulus but can also refer to an integration of at least two stimuli. For instance, stimuli which can only reveal partial information with regard to receptive field location (rotating wedge and contracting ring only convey information on polar angle and eccentricity, respectively) can be integrated into a single stimulus (Wedge-Ring or WR). This integration of information can be performed independent of stimulus presentation. In the case of wedge and ring stimuli each stimulus is presented individually and only after acquisition of the BOLD signal for each is the information integrated. This, however, can be done before or after estimation of pRF parameters. Specifically, integrating the information before estimation is at the level of the signal. It involves the concatenation of signals originating from runs presenting each of these stimuli individually. Estimation is then performed on the concatenated signal. Integrating the information after estimation is at the level of the parameters. In this case, parameters are first estimated separately for each stimulus and the relevant parameter estimates from each stimulus are retained and combined with the retained parameters of the other to generate a full set. Here we considered three stimulus types: bar, wedging pre-estimation (WR_{pre}), and wedge-ring post-estimation (WR_{post}) and supplement the more commonly used orderly presentation of stimuli with random presentation. We include a random presentation sequence since it had previously been shown that a pseudo-random, multi-focal, stimulus improved pRF estimation in the presence of foveal scotomas (Binda et al., 2013). It is as of yet unclear, however, whether this improvement was due to the multi-

focal nature of the stimulus or it being presented in a pseudo-random fashion and whether such choices have any benefit in healthy vision. Regarding continuance, we systematically investigated the effects of including mean luminance periods since it has previously been suggested that their inclusion provides a baseline that allows for improved estimation of large receptive field sizes higher up in the visual hierarchy (Dumoulin & Wandell, 2008). Lastly, data preparation refers to averaging the BOLD signal across stimulus cycles in order to average out random noise in the signal. While averaging should increase the proportion of variance in the BOLD signal that is explained by predictions generated from pRF parameters, it is not clear whether higher proportions of explained variance reflect an improvement in pRF estimation or are simply due to a reduction in the total amount of variance requiring explanation. We aim to distinguish between these possibilities in our evaluation of the effects of averaging.

While we hope to improve the estimation of pRFs and hence their utility to other aspects of vision research, it is apparent that although good estimation of pRFs is surely necessary, it is not always sufficient. That is, it is necessary to assume that pRFs are stable across time in the healthy adult brain a) for situations in which functional localizers and experimental fMRI data are acquired in separate sessions, b) for function based alignment, c) for providing a spatial forward model for computational models, and d) for studying changes due to diseases affecting retinotopic organization (Haak et al., 2014) or due to brain development. Therefore, the second aim of our paper is to test this assumption of stability or temporal consistency.

In order to achieve these aims we study how different choices affect pRF estimation performance in both an empirical as well as in a simulation setting. Simulations allow us to have knowledge of the ground truth regarding pRF size and shape. This renders it possible to investigate in how far choices related to the estimation framework produce correct estimates of each pRF parameter rather than simply explaining variance within the measured BOLD signal.

In summary, we aim to 1) provide a first guideline with respect to choices relating to the estimation framework, and 2) establish the temporal consistency of receptive fields in the healthy adult brain in order to provide a good rationale for vision research relying on knowledge of population receptive fields as well as to further our understanding of them.

Our results with regard to the estimation framework show that a bar stimulus presented in a random (non-continuous) fashion produces the most promising estimates of pRFs, especially if stimulation is interspersed with mean luminance periods. Furthermore, averaging the BOLD signal reduces the amount of variance left unexplained in the signal and might thus convey an advantage in delineating visually responsive voxels. Additionally, for very noisy (high resolution) data sets, our simulations suggest that heavy averaging leads to more accurate pRF parameter estimates while it does not do so if the noise is moderate. Our results with regard to temporal consistency show that pRF estimates originating from two separate measurement sessions separated by a week appear to be very consistent especially with regard to their location and size. The angle of elongation of anisotropic receptive fields, however, is subject

to variations as our simulation results revealed that it is generally more difficult to estimate than the other parameters.

MATERIALS AND METHODS

Participants

Two fMRI measurements separated by 1 week were obtained from three male volunteers. All volunteers were without prior history of psychiatric or neurological illness. The ages in years of subjects one, two, and three were 34, 27, and 35, respectively. All subjects were right handed and had normal or corrected-to-normal visual acuity, were screened, and provided written informed consent prior to scanning.

Ethics Statement and Data Availability

The Ethics Committee of the Faculty of Psychology and Neuroscience at Maastricht University approved this study and the procedures employed therein. The entire empirical data set including both functional and anatomical MR measurements as well as stimulus protocols are available at datadryad.org (doi: 10.5061/dryad.mb8h6).

Stimulus Description

In the present study we used the conventional wedge, ring, and bar stimuli composed of a high-contrast, moving, checkerboard pattern (E. A. DeYoe et al., 1996; Dumoulin & Wandell, 2008; Dumoulin et al., 2003; Lee et al., 2013; Sereno et al., 1995). In contrast to common practice the stimuli did not move continuously across the visual field but in 12 discrete steps and remained at each position for 2 seconds (= 1 volume), covering the entire visual field within 24 seconds. The wedge therefore subtended 30° ($1/12^{\text{th}}$ of the entire 360°) and the width of the rings and bars was $1/12^{\text{th}}$ of the maximum stimulus radius and diameter, respectively. We presented all stimuli in two different modes. The first mode was consecutive with each stimulus position bordering the previous position. For the wedge and ring stimuli this coincides with the phase encoded design first described by Sereno *et al.* (1995). In the first mode the wedge was presented counter clockwise while the ring was contracting and the bar was presented in four different orientations (0 , $\frac{\pi}{4}$, $\frac{\pi}{2}$, and $\frac{3}{4}\pi$ radians) moving in two different directions for each of the orientations (orthogonal to orientation). In the second mode the stimuli were shown at random positions with the exception that for the bar stimulus the positions were only randomized within an orientation and never across orientations. Similar to the original pRF estimation procedure described by Dumoulin and Wandell (Dumoulin & Wandell, 2008) we inserted mean luminance periods during which the participants only saw a zero contrast field. Within each cycle mean luminance fields replaced the stimuli for four consecutive volumes. This amounts to a total of 8 seconds within each cycle wherein subjects were shown

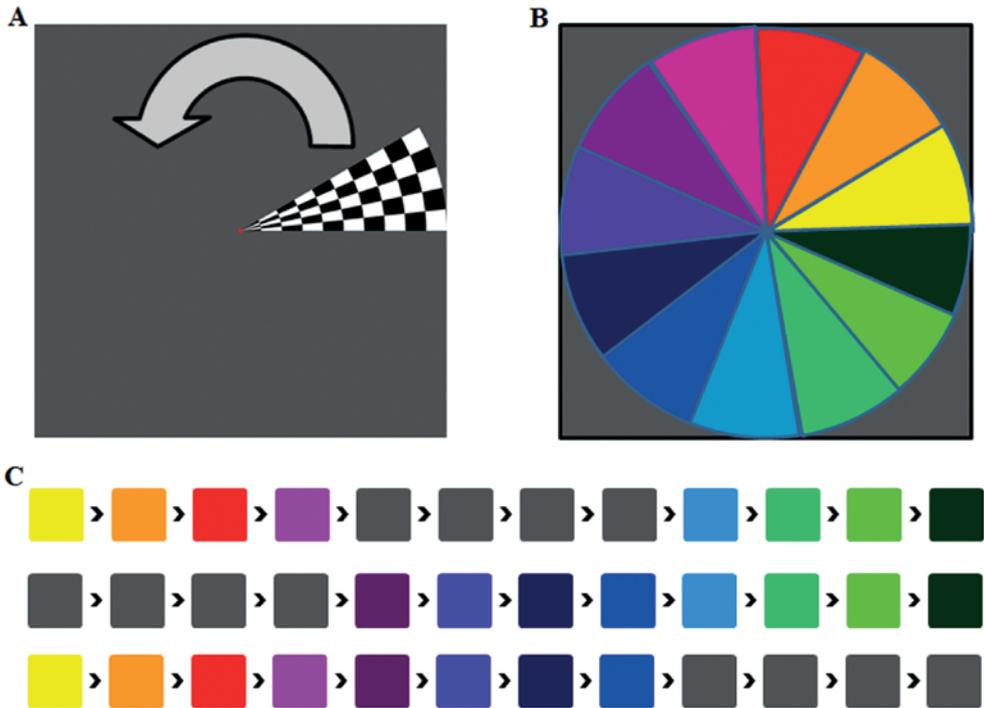


Figure 2.2: Schematic of Stimulus Presentation Scheme. This figure shows a representative schematic of the stimulus presentation scheme including mean luminance periods using the orderly presented wedge stimulus as an example. *Panel A*) Shows a standard wedge aperture located at 0° revealing a checkerboard pattern. The wedge turns counter-clockwise. *Panel B*) Shows a color-code for each of the twelve wedge positions presented during stimulation. *Panel C*) Shows the timecourse of stimulus presentation for three exemplary cycles. The entire stimulation consists of 15 such cycles. The colors of the squares correspond to the color of the wedge position shown at a certain moment in time lasting for two seconds (=1 volume). Grey squares indicate mean luminance periods.

zero contrast rather than a stimulus. The order at which mean luminance fields were inserted was pseudorandom; that is, for each cycle four different positions were replaced such that each position was presented equally often. Figure 2.2 shows a schematic of mean luminance insertion into stimulus cycles using the orderly presented wedge stimulus as an example. For both wedge and ring 15 cycles were presented. For the bar each combination of orientation and direction (4×2) constituted one cycle. There were thus 8 cycles to show each of these combinations and each combination was shown 3 times resulting in a total of 24 cycles. Illustrative visualizations of all stimulus presentations for subject one are provided in the form of videos linked in appendix A1 (all videos are played at fourfold speed of actual stimulation and checkerboards are static rather than flickering).

Stimulus Presentations

The open source stimulus presentation tool StimulGL, developed by authors SG and RG (<https://sites.google.com/site/stimulgl/>) was used for presenting the various visual stimuli to the subjects at a screen resolution of 1680x1050 pixels. The stimuli had a resolution of 1050x1050 pixels. The experiments were performed on a hardware configuration containing a Dell Optiplex 970 computer with a NVIDIA NVS 300 graphics card with OpenGL >2.0 support connected to a Panasonic PT EZ570E wuxga projector. The projected visual stimuli were reflected first with a mirror behind the bore of the magnet and secondly by a mirror above the head coil to the subject. The stimuli filled an area of 40 (width) x 23.5 (height) cm² when projected onto the scanner's frosted screen, corresponding to 30 x 18 degrees visual angle. Given the limit of 18° visual angle in the vertical direction, all stimuli were also limited to 18° in their horizontal extent.

Magnetic resonance imaging

Imaging data were acquired using a 3T Tim Trio scanner equipped with a 32-channel head coil (Siemens Medical Systems, Erlangen, Germany). Anatomical data were collected with a T1-weighted MPRAGE imaging sequence (192 sagittal slices; Repetition Time [TR] = 2250 ms; Echo Time [TE] = 2.17 ms; Flip Angle [FA] = 9°; Field of View [FoV] = 256 x 256 mm²; 1 mm isotropic resolution; GRAPPA = 2). Functional images were acquired using a gradient-echo echo-planar imaging sequence (31 transversal slices; TR = 2000 ms; TE = 30 ms; FA = 77°; FoV = 216 x 216 mm²; 2 mm isotropic resolution; no slice gap; GRAPPA = 2).

Processing of (f)MRI data

All imaging data were analyzed using BrainVoyager QX (v2.6; Brain Innovation, Maastricht, the Netherlands). Anatomical datasets underwent brain extraction, followed by inhomogeneity correction and transformation to ACPC space. Preprocessing of the functional datasets followed standard procedures including slice scan time correction, (rigid body) motion correction, linear trend removal, and temporal high-pass filtering (up to 2 cycles per run). Head motion was minimal (< 1.5 mm translation and < 1.5° rotation in any direction in all runs of every subject) and corrected successfully during post-processing. Due to the use of preparation scans, none of the initial volumes needed to be discarded related to T1 equilibrium effects.

Model Based Analysis

We loosely followed the procedure described by Dumoulin and Wandell (Dumoulin & Wandell, 2008) for estimating the pRF parameters from the time series data using a linear spatio-temporal model of the fMRI response. Specifically, after the BOLD signal measured for each voxel is z-normalized, it is modeled by

$$y_n(t) = \frac{(p(t) - \mu_p)}{\sigma_p} + e \quad (2.1)$$

where $p(t)$ is the predicted BOLD signal, e represents additive Gaussian noise, and μ_p and σ_p respectively represent the mean and standard deviation of the predicted BOLD signal. In accordance with Dumoulin and Wandell (Dumoulin & Wandell, 2008) the prediction $p(t)$ was calculated from a parameterized model of the underlying modeled neuronal population receptive field and the stimulus. We estimated the pRF model by finding those model parameters that best fit the data. Here, we modeled the pRF as an anisotropic two-dimensional Gaussian. We used an anisotropic model for reasons of completeness as it allows us to investigate the aptitude of each stimulus in identifying all pRF properties. Additionally, previous research had shown that an anisotropic model is beneficial for the estimation of pRF location at eccentricities close to the outer boundaries of the stimulated visual field segment (Lee et al., 2013). We did, however, repeat the analyses with an isotropic pRF model and obtained the same results with regard to stimulus choice. The anisotropic Gaussian is defined by five parameters $(x_0, y_0, \sigma_x, \sigma_y, \theta)$,

$$g(x, y) = e^{-\left(a(x-x_0)^2 + 2b(x-x_0)(y-y_0) + c(y-y_0)^2\right)} \quad (2.2)$$

And

$$a = \frac{\cos^2\theta}{2\sigma_x^2} + \frac{\sin^2\theta}{2\sigma_y^2} \quad (2.3)$$

$$b = -\frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2} \quad (2.4)$$

$$c = \frac{\sin^2\theta}{2\sigma_x^2} + \frac{\cos^2\theta}{2\sigma_y^2} \quad (2.5)$$

where (x_0, y_0) is the center of the Gaussian, σ_x and σ_y is the spread in the X and Y directions, respectively, and θ is the rotation of the Gaussian. From this receptive field model and an effective stimulus, $s(x, y, t)$, which is a binary indicator function that marks the position of the stimulus aperture at each time, we calculated the BOLD signal in two steps. First, we calculated the overlap between the effective stimulus and the model pRF at each point in time:

$$r(t) = \sum_{x,y} s(x, y, t) g(x, y) \quad (2.6)$$

Second, we convolved this overlap with a two-gamma hemodynamic response function h :

$$p(t) = r(t) * h(t) \quad (2.7)$$

In order to find the optimal pRF parameters we performed a grid search. The grid consisted of a set of 36000 plausible parameter configurations. Specifically, we used a 20 by 20 grid of central positions ranging from -9° to $+9^\circ$ (measured from fixation), a set of 10 exponentially increasing spreads from 0.25 to 10 [we expected most RFs to be comparatively small and only a few to be big], two ratios of spread in the X and Y direction ($\sigma_x/\sigma_y = 1:1$ and $2:1$), one angle of rotation if the ratio is 1:1, and eight angles of rotation if the ratio is 2:1. The search space was traversed for all voxels simultaneously using a graphical processing unit (GPU) accelerated parallelization procedure. Each voxel was assigned the parameter set which minimizes the normalized root-mean-square error (NRMSE) between its Z-normalized predicted and Z-normalized measured BOLD signal:

$$NRMSE = \sqrt{\frac{\sum_i (y_n(t) - p_n(t))^2}{T}} / 2.35 \quad (2.8)$$

where T is the number of time points, 2.35 is the full width at half maximum of the Z-distribution, and the subscript n indicates that the data is normalized to Z-scores. Additionally, a fit value was retained for each voxel given by $1 - NRMSE$.

Information Integration

With respect to pRF location the wedge and ring stimuli can only provide adequate information on either polar angle or eccentricity, respectively. We, therefore, only considered combinations of these stimuli in estimating pRF parameters. Specifically, we considered two modes of combination. For the first mode the stimuli were combined before pRF estimation by concatenation of BOLD signals obtained for each stimulus individually. The pRF parameters were then estimated from this concatenated signal (we refer to this stimulus setting as WR_{pre}). For the second mode the pRF parameters were first estimated for each stimulus individually. Afterwards, the Cartesian coordinates obtained for the wedge were transformed to angular values while those of the ring were transformed to radii. Next, the combined location was obtained by transforming these angle and radius values back to the Cartesian coordinate system. Finally, the remaining parameter values were averaged¹ across the two stimuli (we refer to this stimulus setting as WR_{post}). The bar stimulus (hereafter referred to as bar) arguably provides adequate information on all pRF parameters.

Voxel Selection

In the present study, voxels were treated as subjects in the statistical analyses. Since it is only meaningful to include those voxels that encode visual information it was necessary to

¹ To obtain the average of the rotation angle we use the formula for the mean of circular quantities

separate visual from non-visual voxels for each subject and scan session. This was done in two steps. First, voxels were sorted into two clusters by performing k-means clustering on their fit values for each stimulus configuration (i.e. experimental run) individually. This led to a cluster of visual voxels per experimental run. In the second step only voxels which were consistently placed into the visual cluster for all runs were retained in the analyses. In the first session 2655, 2669, and 1969 voxels were retained for the three subjects, respectively. In the second session 2592, 2764, and 1673 voxels were retained for subjects one, two, and three respectively. In order to check the validity of this data-driven voxel selection procedure we changed the first step to labeling all voxels visual whose Fit values were larger than the 95th, 97.5th, and 99th percentile. After the second step, the final voxel selection for all three percentiles was identical with that derived using the original procedure.

Simulated Data

In addition to the empirical data set, we simulated fMRI data from known receptive fields. Specifically, the simulated data was given by

$$y_s(t) = p(t) + e \quad (2.9)$$

where $p(t)$ is the predicted BOLD signal, and e is an error term obtained from the residuals of the pRF estimation of subject 1. The BOLD signal $p(t)$ was given by applying equations 2.2 through 2.7 using a pre-defined set of receptive field parameters, the effective stimulus settings for the wedge, ring, and bar stimuli we previously used in the experimental setup of subject 2, and a prototypical hemodynamic response obtained from BOLD signal of subject 3. The set of pre-defined pRF parameters consists of 36000 parameter configurations each of which corresponds exactly to a point on the search grid. For this set of simulated data all assumptions regarding the pRF model were met while at the same time we searched the parameter space optimally. Therefore, all deviations of estimated pRFs with respect to the true underlying pRFs reflect on the aptness of the stimulus rather than on model choice and parameter estimation procedures.

Metrics

In order to evaluate the estimation framework as well as the extent to which population receptive fields are consistent over time we calculated a set of metrics. The first of these, exclusively used to evaluate the estimation framework for empirical data, was explained variance. Specifically, we quantified the variance of the measured BOLD signal y_m that was explained by the predicted BOLD signal y_p as the squared Pearson correlation between y_m and y_p .

The second metric, used to evaluate the estimation framework for simulated data as well as to measure temporal consistency of pRFs, was the similarity S between model parameters of two receptive fields² a and b given by

$$S_{ab} = 1 - \sqrt{\frac{\left(\frac{X_a - X_b}{X_{max} - X_{min}}\right)^2 + \left(\frac{Y_a - Y_b}{Y_{max} - Y_{min}}\right)^2 + \left(\frac{\sigma_{x,a} - \sigma_{x,b}}{\sigma_{x,max} - \sigma_{x,min}}\right)^2 + \left(\frac{\sigma_{y,a} - \sigma_{y,b}}{\sigma_{y,max} - \sigma_{y,min}}\right)^2 + \left(\frac{\arg(\exp(2(\theta_a - \theta_b)i))}{\pi}\right)^2}{5}} \quad (2.10)$$

The differences between pRFs were normalized for each parameter with respect to their range in order to account for the different scales of the parameters and hence to obtain a similarity metric for which every parameter carries equal weight. Additionally, for angles of elongation we calculated an angular difference since an angle of 0 and an angle of π are identical with respect to pRF shape. The factor of two in the angular difference accounts for the fact that the range from 0 to π needs to be stretched around the entire circle for the formula to apply.

Finally, another metric of temporal consistency of pRFs was the correlation of Polar Angle and Eccentricity maps across sessions. For Eccentricity maps this was simply the Pearson correlation coefficient. Since Polar Angle represents circular values, we used the circular correlation coefficient r_{circ} proposed by Fisher and Lee (Fisher & Lee, 1983) for Polar Angle maps:

$$r_{circ} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sin(\alpha_i - \alpha_j) \sin(\beta_i - \beta_j)}{\sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sin^2(\alpha_i - \alpha_j) \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sin^2(\beta_i - \beta_j)}} \quad (2.11)$$

where n is the number of data points and α and β are angles.

Analyses and Information Aggregation

To establish which choices with regard to the estimation framework are most optimal a ranking procedure based on the Tideman method was employed (Tideman, 1987). For example, in order to establish which combination of stimulus and mode of presentation should be used, all possible combinations of these choices were pairwise compared to each other with respect to the relevant outcome metric. This was done per subject per session such that each subject casted two votes (one for each session) as to which member of a pair is preferable. The votes were Hedges' g (Hedges, Olkin, & Statistiker, 1985) values which provide a standardized distance between means. Specifically, Hedges' g is given by

² These receptive fields always refer to the same voxel but to two obtained parameter sets, for instance, using different stimuli or from different sessions.

$$g = \frac{\bar{x}_1 - \bar{x}_2}{s} \quad (2.12)$$

where \bar{x}_1 and \bar{x}_2 are the sample means and s the pooled standard deviation given by

$$S = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (2.13)$$

with s_1 and s_2 being the standard deviations and n_1 and n_2 the sample sizes of the two samples, respectively. We created 95% percent confidence intervals around the Hedges' g values and considered a member of a pair to be preferable to the other only if zero did not fall within this confidence interval. Hence, a vote for one member of a pair was cast if it produced a significantly higher value on the outcome metric than its rival and the vote was weighted by the actual value of Hedges' g to ensure that strong differences between members carry more weight than weak, albeit significant, differences. The pairwise Hedges' g values of all subjects and sessions were added up to obtain a matrix of pairwise voting results. The Tideman method (Tideman, 1987) was then applied to this matrix of voting results in three steps. 1) Tally: for each pair (A,B) in the matrix the sum of Hedges' g values for A winning over B was compared to the sum of Hedges' g values for B winning over A. The larger of the two sums represented a majority and hence determined the pair's winner while the smaller of the two sums represented a minority. 2) Sort: the pairs were ordered based on their majorities with larger majorities preceding smaller ones. In case the winners of two pairs had equal majorities, the pair with the smaller minority was ranked first. 3) Lock: a directed graph in the form of an adjacency matrix was constructed from these pairs. In an adjacency matrix each present connection (edge) between a pair of candidates (vertices) is represented by 1 while an absent connection is represented by 0. Here, edges were drawn from a pair's winner to a pair's loser with circular connections omitted. The indegree (number of incoming connections) of each vertex determines its rank with a lower indegree corresponding to a higher rank. Consequently the source of the graph (the vertex with an indegree of zero) is the overall winner.

This procedure was employed for both empirical as well as simulation results. Hedges' g measures of effect sizes rather than pairwise t-tests were used for two reasons. Firstly, all statistical tests have tremendous power since voxels constitute our sample and the associated sample size renders even negligible effects highly significant. This is especially true for analyses performed on empirical data. Secondly, effect sizes are more informative with respect to best practice in an experimental setup as they provide a straightforward mode of comparison (Steiger, 2003).

RESULTS

Optimization of the Estimation Framework

Stimulus Type and Stimulus Presentation

The first ranking was performed to identify which combination of stimulus type and mode of presentation yields a set of pRF parameters which can predict BOLD signals obtained from different experimental setups measured during the same scanning session. This is to ensure that an estimation framework does not yield stimulus-specific population receptive field estimates. From each set of pRF parameters obtained from a specific combination of stimulus and mode of presentation we predicted BOLD signals for those experimental runs during which another stimulus was presented. Subsequently, we calculated the proportion of variance observed in the measured BOLD signals that was explained by those predictions. The explained variance values obtained for every predicted run were averaged to obtain a single value per voxel. For instance, if parameters were obtained using the bar stimulus, BOLD signals for both the wedge and ring (using both presentation modes) were predicted from the parameter set of

Table 2.1: Ranking for within-session explanatory power

A. Voting results for within-session explanatory power

		WR _{pre}		WR _{post}		Bar	
		orderly	random	orderly	random	orderly	random
WR _{pre}	orderly		0.5	0.4	0.3	0.2	0.2
	random	0.6		1.1	0.8	0.4	0.4
WR _{post}	orderly	0.0	0.5		0.0	0.2	0.2
	random	0.3	0.7	0.5		0.2	0.2
Bar	orderly	1.1	1.2	1.6	1.2		0.0
	random	1.6	1.6	2.0	1.5	0.3	

B. Pairwise comparisons for within-session explanatory power

Pair	Order
WR _{pre} , orderly ($g = 0.5$) vs. WR _{pre} , random ($g = 0.6$)	11
WR _{pre} , orderly ($g = 0.4$) vs. WR _{post} , orderly ($g = 0$)	13
WR _{pre} , orderly ($g = 0.3$) vs. WR _{post} , random ($g = 0.3$)	15
WR _{pre} , orderly ($g = 0.2$) vs. Bar, orderly ($g = 1.1$)	8
WR _{pre} , orderly ($g = 0.2$) vs. Bar, random ($g = 1.6$)	2
WR _{pre} , random ($g = 1.1$) vs. WR _{post} , orderly ($g = 0.5$)	9
WR _{pre} , random ($g = 0.8$) vs. WR _{post} , random ($g = 0.7$)	10
WR _{pre} , random ($g = 0.4$) vs. Bar, orderly ($g = 1.2$)	7
WR _{pre} , random ($g = 0.4$) vs. Bar, random ($g = 1.6$)	4
WR _{post} , orderly ($g = 0$) vs. WR _{post} , random ($g = 0.5$)	12
WR _{post} , orderly ($g = 0.2$) vs. Bar, orderly ($g = 1.6$)	3
WR _{post} , orderly ($g = 0.2$) vs. Bar, random ($g = 2$)	1
WR _{post} , random ($g = 0.2$) vs. Bar, orderly ($g = 1.2$)	6
WR _{post} , random ($g = 0.2$) vs. Bar, random ($g = 1.5$)	5
Bar, orderly ($g = 0$) vs. Bar, random ($g = 0.3$)	14

C. Directed graph for within-session explanatory power

		WR _{pre}		WR _{post}		Bar	
		orderly	random	orderly	random	orderly	random
WR _{pre}	orderly	0	0	1	0	0	0
	random	1	0	1	1	0	0
WR _{post}	orderly	0	0	0	0	0	0
	random	0	0	1	0	0	0
Bar	orderly	1	1	1	1	0	0
	random	1	1	1	1	1	0
Indegree		3	2	5	3	1	0

A) Tally: Each cell represents the sum of hedges' g values for the pairwise comparison of row over column obtained for each combination of subject and session. Since three subjects were tested in two sessions (each subject casted two votes), there were six votes in total. Only significant results were summed, i.e. those hedges' g values whose confidence intervals did not include zero. If zero lay within the confidence interval the vote was counted as indifferent between the two options. **B) Sort:** The first column shows the comparison of each pair including the hedges' g values of each winning over the other. The winner of each pair is highlighted in bold. The second column shows the order in which pairs were locked based on majorities starting with the largest. **C) Lock:** The graph depicts binary edges leading from row to column as well as the sum total of incoming edges for each vertex (stimulus-presentation combination).

each voxel and the explained variance values obtained for the resulting predictions pertaining to wedge and ring were averaged. From these explained variances we derived votes in the form of Hedges' g measures of effect size (see methods for details). Table 2.1a shows the voting results for all pairwise comparisons of stimulus-presentation combinations with every voting result indicating the sum of Hedges' g values for the one-sided comparisons of row larger than column.

Applying the Tideman method (Tideman, 1987) to these voting results, we considered each pairwise comparison in turn and determined its winner (tally). Subsequently, we ordered the pairs from largest to smallest majority (sort). Table 2.1b shows a summary of these steps. Finally, the results were locked in a directed graph (lock). The adjacency matrix of this graph as well as each pair's indegree is represented in table 2.1c. The randomly presented bar was the source of the graph and thus constituted the top of the hierarchy. The orderly presented bar was ranked second with an indegree of 1, followed by randomly presented WR_{pre} with an indegree of 2. The fourth rank was shared by orderly presented WR_{pre} and randomly presented WR_{post} which both had an indegree of 3. Finally, orderly presented WR_{post} was ranked lowest with an indegree of 5.

Based on the rankings of the combinations of stimulus type and mode of presentation we derived separate matrices of voting results for stimulus type and mode of presentation. To determine the votes of A winning over B , where A and B are different stimulus types, we examined how often the orderly variant of A was ranked above both variants of B and how often the random variant of A was ranked above both variants of B . For instance, the orderly variant of WR_{pre} was ranked only above the orderly variant of WR_{post} while the random variant of WR_{pre} was ranked above both variants of WR_{post} . This yields a total of three instances where WR_{pre} was ranked above WR_{post} . Table 2.2a shows the voting results obtained accordingly for all stimulus pairs.

Similarly, to determine the votes of A winning over B , where A and B are different modes of presentation, we examined how often each stimulus presented in variant A was ranked above all stimuli presented in variant B . Table 2.2b shows the voting results for the comparison of orderly and random presentations. Since the bar was ranked consistently above the other stimuli, it constituted the Condorcet winner³ in the context of predicting BOLD activity obtained from other retinotopic mapping stimuli. Since in the majority of rankings the random presentation variant was preferable to the orderly presentation variant, it constituted the Condorcet winner in the context of predicting BOLD activity of other stimuli.

3 A candidate in an election is considered the Condorcet winner if that candidate wins by majority rule against each other candidate. This condition to determine a winner was developed by the Marquis de Condorcet in 1785 and described in his work (Condorcet, 1785), *An essay on the application of probability theory to plurality decision making: An election between three candidates*.

Table 2.2: Voting results for within-session explanatory power split for stimulus and mode of presentation

A. Stimulus type				B. Mode of presentation		
	WR_{pre}	WR_{post}	Bar		Orderly	Random
WR_{pre}		3	0	Orderly		2
WR_{post}	0		0	Random	6	
Bar	4	4				

A) Stimulus type: Each cell represents the sum of rankings where the row stimulus was ranked above the column stimulus. **B) Mode of presentation:** Each cell represents the sum of rankings where the mode of presentation given by the row was ranked above the mode of presentation given by the column.

The second ranking was performed in order to identify which combination of stimulus and mode of presentation yields a set of pRF parameters from which a BOLD signal measured at a different point in time can be predicted most accurately. To this end we again calculated the proportion of variance observed in the measured BOLD signal that a model BOLD signal based on pRF parameters from a specific combination can explain. In contrast to the previous analysis, however, the pRF parameters were obtained in one session while the measured BOLD signal was taken from a second session.

Table 2.3a shows the voting results for the variance explained across sessions (tally). By applying the Tideman method (Tideman, 1987) we obtained the winners of each pairwise comparison as well as their orderings (sort). The comparisons, winners, and orderings are given in table 2.4b. Subsequently, we obtained the ranking of the combinations by locking the results in a directed graph (lock; see table 2.3c). The randomly presented bar was the source of the graph and thus constituted the top of the hierarchy. The orderly presented bar was ranked second with an indegree of 1, followed by randomly presented WR_{pre} with an indegree of 2. WR_{post} presented randomly was ranked fourth with an indegree of 3. The fifth rank was occupied by orderly presented WR_{pre} with an indegree of 4. Finally, orderly presented WR_{post} was ranked lowest with an indegree of 5.

Based on the rankings of the combinations of stimulus type and stimulus presentation we derived matrices of voting results for type and presentation, respectively. Table 2.4a shows the voting results with regard to stimulus. Since the bar was ranked consistently above the other stimuli, it constituted the Condorcet winner in the context of predicting BOLD activity across sessions.

With regard to stimulus presentation, table 2.4b shows the voting results for the comparison of orderly and random presentations. Since for the majority of rankings the random presentation variant was preferable to the orderly presentation variant, it constituted the Condorcet winner in the context of predicting BOLD activity across sessions.

Table 2.3: Ranking for across-session explanatory power

A. Voting results for across-session explanatory power

		WR _{pre}		WR _{post}		Bar	
		orderly	random	orderly	random	orderly	random
WR _{pre}	orderly		0.4	0.3	0.2	0.3	0.1
	random	0.5		0.9	0.7	0.5	0.3
WR _{post}	orderly	0.0	0.4		0.0	0.2	0.1
	random	0.3	0.5	0.4		0.4	0.2
Bar	orderly	0.9	1.0	1.3	1.0		0.0
	random	1.2	1.2	1.5	1.3	0.3	

B. Pairwise comparisons for across-session explanatory power

Pair	Order
WR _{pre} , orderly (g = 0.4) vs. WR _{pre} , random (g = 0.5)	11
WR _{pre} , orderly (g = 0.3) vs. WR _{post} , orderly (g = 0)	13
WR _{pre} , orderly (g = 0.2) vs. WR _{post} , random (g = 0.3)	15
WR _{pre} , orderly (g = 0.3) vs. Bar, orderly (g = 0.9)	8
WR _{pre} , orderly (g = 0.1) vs. Bar, random (g = 1.2)	4
WR _{pre} , random (g = 0.9) vs. WR _{post} , orderly (g = 0.4)	9
WR _{pre} , random (g = 0.7) vs. WR _{post} , random (g = 0.5)	10
WR _{pre} , random (g = 0.5) vs. Bar, orderly (g = 1)	7
WR _{pre} , random (g = 0.3) vs. Bar, random (g = 1.2)	5
WR _{post} , orderly (g = 0) vs. WR _{post} , random (g = 0.4)	12
WR _{post} , orderly (g = 0.2) vs. Bar, orderly (g = 1.3)	2
WR _{post} , orderly (g = 0.1) vs. Bar, random (g = 1.5)	1
WR _{post} , random (g = 0.4) vs. Bar, orderly (g = 1)	6
WR _{post} , random (g = 0.2) vs. Bar, random (g = 1.3)	3
Bar, orderly (g = 0) vs. Bar, random (g = 0.3)	14

C. Directed graph for across-session explanatory power

		WR _{pre}		WR _{post}		Bar	
		orderly	random	orderly	random	orderly	random
WR _{pre}	orderly	0	0	1	0	0	0
	random	1	0	1	1	0	0
WR _{post}	orderly	0	0	0	0	0	0
	random	1	0	1	0	0	0
Bar	orderly	1	1	1	1	0	0
	random	1	1	1	1	1	0
Indegree		4	2	5	3	1	0

A) Tally: Each cell represents the sum of hedges' g values for the pairwise comparison of row over column obtained for each combination of subject and session. Since three subjects were tested in two sessions (each subject casted two votes), there were six votes in total. Only significant results were summed, i.e. those hedges' g values whose confidence intervals did not include zero. If zero lay within the confidence interval the vote was counted as indifferent between the two options. **B) Sort:** The first column shows the comparison of each pair including the hedges' g values of each winning over the other. The winner of each pair is highlighted in bold. The second column shows the order in which pairs were locked based on majorities starting with the largest. **C) Lock:** The graph depicts binary edges leading from row to column as well as the sum total of incoming edges for each vertex (stimulus-presentation combination).

Table 2.4: Voting results for across-session explanatory power split for stimulus and mode of presentation

A. Stimulus type

	WR _{pre}	WR _{post}	Bar
WR _{pre}		3	0
WR _{post}	1		0
Bar	4	4	

B. Mode of presentation

	Orderly	Random
Orderly		2
Random	7	

A) Stimulus type: Each cell represents the sum of rankings where the row stimulus was ranked above the column stimulus. **B) Mode of presentation:** Each cell represents the sum of rankings where the mode of presentation given by the row was ranked above the mode of presentation given by the column.

Stimulus Type and Continuance

The third ranking was performed for simulated data with the aim to investigate which stimulus type either interspersed by mean luminance (ML) periods or not (−ML) most faithfully recovers known population receptive field parameters. To this end we obtained the similarity between original and recovered pRF parameters. Since neither attention nor lateral connectivity was

modeled in our simulations all stimuli were presented in an orderly fashion only. However, each stimulus was presented both with and without mean luminance periods in two separate runs. The presence of mean luminance periods has been argued to improve the estimation of receptive field parameters, especially if receptive fields are large (Dumoulin & Wandell, 2008). As with the empirical results rankings are performed on the Hedges' g values obtained for the pairwise comparison of all combinations of stimulus and mean luminance. Table 2.5a shows the resulting votes.

Applying the Tideman method (Tideman, 1987) we obtained the winners of each pairwise comparison as well as their orderings (tally). The comparisons, winners, and orderings are given in table 2.5b. Subsequently, we obtained the ranking of the combinations by locking the results in a directed graph (lock; see table 2.5c). The resulting graph had two sources since both versions of the bar, that is presented interspersed and not interspersed by periods of mean luminance, had only outgoing but no incoming connections. With an indegree of 1 WR_{pre} presented interspersed with mean luminance periods was ranked second. Finally, all remaining combinations of stimulus and presence or absence of mean luminance periods had an indegree of 3 and thus constitute equally undesirable choices.

Table 2.5: Ranking for pRF estimate Similarity

A. Voting results for pRF estimate Similarity							
		WR_{pre}		WR_{post}		Bar	
		ML	~ML	ML	~ML	ML	~ML
WR_{pre}	ML		2.1	1.6	1.9	0.0	0.0
	~ML	0.0		0.0	0.0	0.0	0.0
WR_{post}	ML	0.0	0.0		0.0	0.0	0.0
	~ML	0.0	0.0	0.0		0.0	0.0
Bar	ML	1.0	3.1	2.4	2.8		0.0
	~ML	0.0	2.6	2.0	2.3	0.0	

B. Pairwise comparisons for pRF estimate Similarity		
Pair	Order	
WR_{pre}, ML (g = 2.1) vs. WR_{pre}, ~ML (g = 0)	6	
WR_{pre}, ML (g = 1.6) vs. WR_{post}, ML (g = 0)	8	
WR_{pre}, ML (g = 1.9) vs. WR_{post}, ~ML (g = 0)	7	
WR_{pre} , ML (g = 0) vs. Bar , ML (g = 1)	9	
WR_{pre} , ML (g = 0) vs. Bar , ~ML (g = 0)	10	
WR_{pre} , ~ML (g = 0) vs. WR_{post} , ML (g = 0)	11	
WR_{pre} , ~ML (g = 0) vs. WR_{post} , ~ML (g = 0)	12	
WR_{pre} , ~ML (g = 0) vs. Bar , ML (g = 3.1)	1	
WR_{pre} , ~ML (g = 0) vs. Bar , ~ML (g = 2.6)	3	
WR_{post} , ML (g = 0) vs. WR_{post} , ~ML (g = 0)	13	
WR_{post} , ML (g = 0) vs. Bar , ML (g = 2.4)	4	
WR_{post} , ML (g = 0) vs. Bar , ~ML (g = 2)	14	
WR_{post} , ~ML (g = 0) vs. Bar , ML (g = 2.8)	2	
WR_{post} , ~ML (g = 0) vs. Bar , ~ML (g = 2.3)	5	
Bar , ML (g = 0) vs. Bar , ~ML (g = 0)	15	

C. Directed graph for pRF estimate Similarity							
		WR_{pre}		WR_{post}		Bar	
		ML	~ML	ML	~ML	ML	~ML
WR_{pre}	ML	0	1	1	1	0	0
	~ML	0	0	0	0	0	0
WR_{post}	ML	0	0	0	0	0	0
	~ML	0	0	0	0	0	0
Bar	ML	1	1	1	1	0	0
	~ML	0	1	1	1	0	0
Indegree		1	3	3	3	0	0

A) Tally: Each cell represents the sum of hedges' g values for the pairwise comparison of row over column obtained for each combination of subject and session. Since three subjects were tested in two sessions (each subject casted two votes), there were six votes in total. Only significant results were summed, i.e. those hedges' g values whose confidence intervals did not include zero. If zero lay within the confidence interval the vote was counted as indifferent between the two options. **B) Sort:** The first column shows the comparison of each pair including the hedges' g values of each winning over the other. The winner of each pair is highlighted in bold. The second column shows the order in which pairs were locked based on majorities starting with the largest. **C) Lock:** The graph depicts binary edges leading from row to column as well as the sum total of incoming edges for each vertex (stimulus-presentation combination).

Table 2.6: Voting results for pRF estimate Similarity split for stimulus and mode of presentation

A. Stimulus type				B. Continuance		
	WR _{pre}	WR _{post}	Bar		ML	¬ML
WR _{pre}		2	0	ML		4
WR _{post}	0		0	¬ML	2	
Bar	4	4				

A) Stimulus type: Each cell represents the sum of rankings where the row stimulus was ranked above the column stimulus. **B) Mode of presentation:** Each cell represents the sum of rankings where the mode of presentation given by the row was ranked above the mode of presentation given by the column.

Based on the rankings of the combinations of stimulus type and continuance we derived matrices of voting results for type and continuance, respectively. Table 2.6a shows the voting results with regard to stimulus. Since the bar was ranked consistently above the other stimuli, it constituted the Condorcet winner for the simulations.

With regard to continuance, table 2.6b shows the voting results for the comparison of presentations with and without interspersed mean luminance periods. Since for the majority of rankings the presence of mean luminance periods was preferable to their absence, the inclusion of mean luminance periods constituted the Condorcet winner for the simulations.

The beneficial effect of mean luminance periods is supposed to manifest itself mainly for later visual areas. For this reason, we repeated the current ranking for V1 and V4 type receptive fields separately. To create V1 and V4 type receptive fields we used two different sets of receptive field sizes while keeping the remaining parameters constant. The sizes were a linear function of the eccentricity with biologically realistic slopes. Specifically, the slope for V4 (.84) was much steeper than the slope for V1 (.21) (Freeman & Simoncelli, 2011). We concentrate here only on the main findings of these additional rankings and provide detailed analysis in appendices A2 and A3. As expected for V1 type receptive fields, neither stimulus choice nor the inclusion of mean luminance periods had strong effects on pRF estimation performance. Nonetheless, the effect of mean luminance is opposite to what has been proposed for late visual areas. Namely, omission of mean luminance periods leads to slightly better pRF estimates than their inclusion. With regard to stimulus choice, WR_{post} and bar both ranked above WR_{pre} but were indistinguishable among themselves (for a detailed analysis see appendix A2). For V4 the general findings of our third ranking reappear. With regard to stimuli, the bar was superior to both WR_{pre} and WR_{post} with WR_{pre}, in turn, being superior to WR_{post}. With regard to the mean luminance periods, it was clearly beneficial to include such periods in pRF estimation (for a detailed analysis see appendix A3). These findings support the notion that mean luminance periods aid the estimation of visual areas with large receptive field sizes.

Data Preparation

A final consideration pertains to the advantages of performing pRF estimation on BOLD signal averaged across several stimulus cycles. To this end we compared the explained variance as well as Similarity among pRF parameters for estimated and known underlying population receptive fields for averaged and non-averaged simulated BOLD signals. The simulated data came from a procedure using a bar stimulus presented in an orderly fashion without mean luminance gaps. This estimation framework was chosen since it represents straightforward matches of stimulus positions from one cycle to the next. Three different conditions were considered. In the first, three repetitions of the 8 orientation (4) and direction (2) combinations were used to generate a non-averaged simulated BOLD signal. In the second, the BOLD signal resulting from these three repetitions was averaged. Finally, in the third condition, BOLD

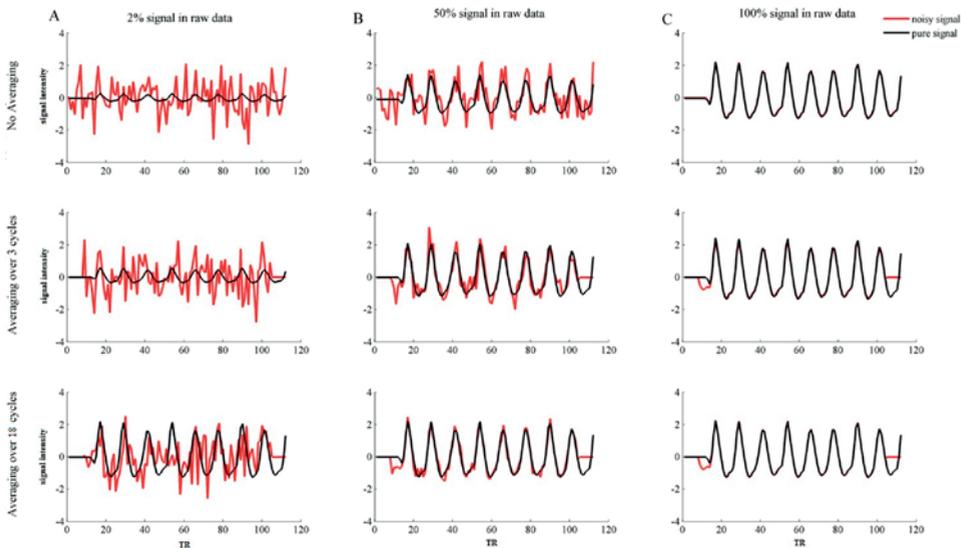


Figure 2.3: Averaged and non-averaged BOLD signals for different noise levels. This figure shows both averaged and non-averaged versions of BOLD signals for three levels of noise for one simulated voxel. The red line indicates the overall BOLD signal resulting from stimulation as well as additive noise. Note that the baseline period (8 TRs) before and after the stimulation period are set to zero. The black line indicates the signal purely as a function of stimulation contained within the larger BOLD signal. *Panel A*) shows non-averaged signal (upper row), signal averaged over 3 cycles (middle row), and signal averaged over 20 cycles (lower row) for raw (i.e. not averaged) BOLD of which 2 percent is stimulus-related signal. More extensive averaging reduces noise to a greater extent and hence increases the stimulus-related variation in the overall signal. Specifically, only for signal averaged over 20 cycles does the overall signal resemble the purely stimulation related signal. *Panel B*) shows non-averaged signal (upper row), signal averaged over 3 cycles (middle row), and signal averaged over 20 cycles (lower row) for raw BOLD of which 50 percent is stimulus-related signal. More extensive averaging reduces noise to a greater extent and hence increases the stimulus-related variation in the overall signal. Nonetheless, even for non-averaged data, the overall signal bears resemblance to the purely stimulus-related signal. *Panel C*) shows non-averaged signal (upper row), signal averaged over 3 cycles (middle row), and signal averaged over 20 cycles (lower row) for raw BOLD of which 100 percent is stimulus-related signal. Since there is no noise in the data, averaging has no effect.

signal was generated from 18 stimulus repetitions and subsequently averaged. Additionally, we parametrically varied the amount of additive noise so as to investigate the advantage that averaging conveys for hardly noisy to very noisy data sets. Figure 2.3 shows both averaged and non-averaged BOLD signals for different noise levels. The red lines indicate the signal including noise while the black lines indicate the pure signal resulting from stimulation contained within the noisy signal. Especially for noisy signals, i.e. where signal resulting from stimulation constitutes merely 2 percent of the resulting BOLD signal (figure 2.3a), averaging lead to profound increases in stimulation related variance contained in the overall noisy signal. Population receptive field parameters were estimated from averaged as well as non-averaged signals for each level of noise.

Figure 2.4 shows the variance explained by a predicted BOLD signal where the prediction is based on pRF parameters estimated from non-averaged signals (upper row), signals averaged over 3 cycles (middle row), and signals averaged over 18 cycles (lower row), respectively. As expected from the higher stimulation-related signal content for averaged signals, averaging had a tremendous benefit on the explained variance. Additionally, more averaging resulted in larger proportions of explained variance.

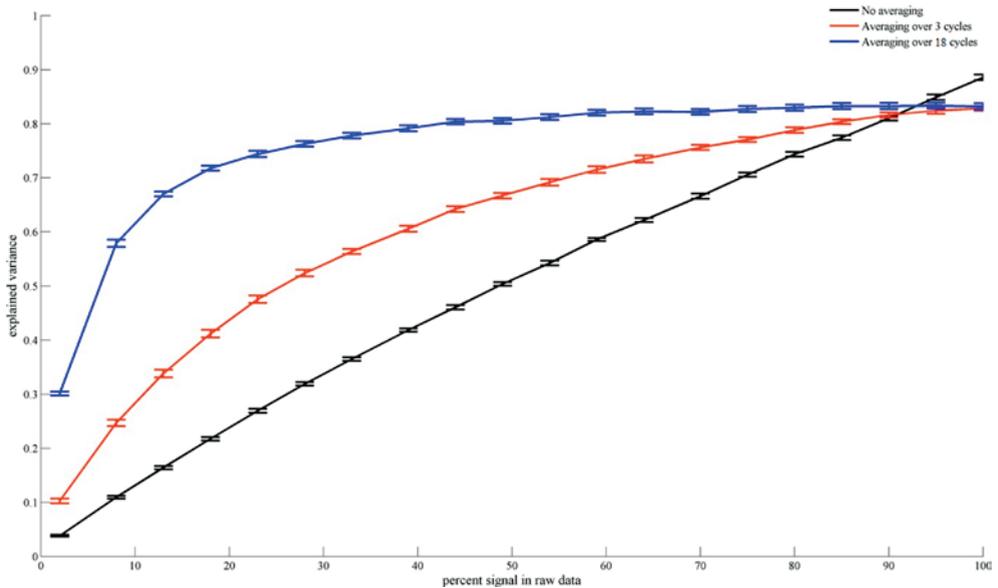


Figure 2.4: Explained variance within averaged and non-averaged BOLD signals for different noise levels.

This figure shows the amount of variance explained for averaged and non-averaged BOLD signals based on pRF parameters estimated from these signals as a function of the percent stimulus-related signal in the raw data. The explained variance for non-averaged signals (black line) scales linearly with the percentage of stimulus-related signal in the raw data. That is, the estimated pRF parameters explain as much variance of the overall signal as is actually due to stimulation. For averaged signals, the amount of variance explained both starts off higher and reaches a plateau faster than for the non-averaged signal. This effect is stronger for signal averaged across 20 cycles (blue line) than for signal averaged across 3 cycles (red line).

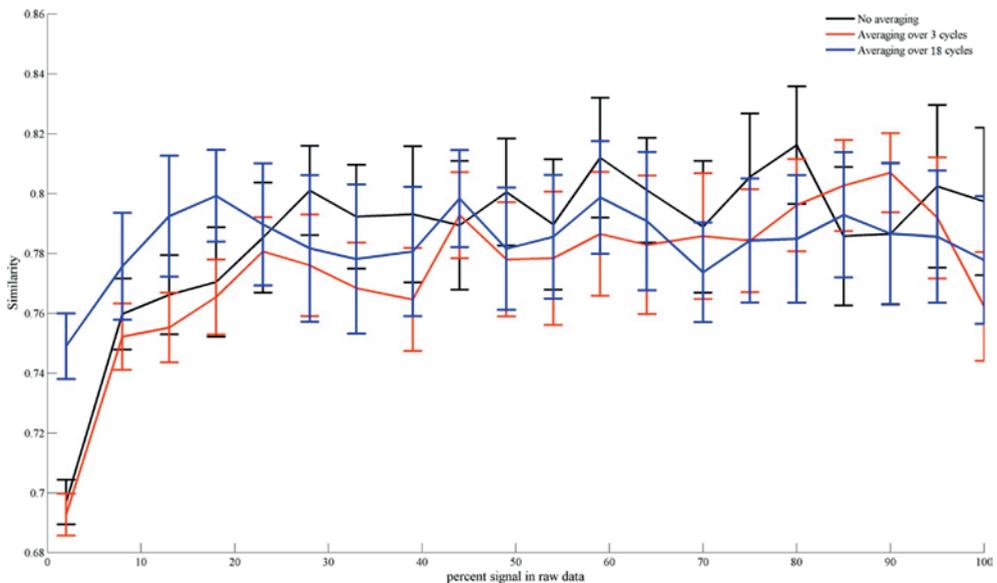


Figure 2.5: Similarity between known and estimated pRF parameters for averaged and non-averaged BOLD signals at different noise levels. The figure shows the Similarity between known and estimated pRF parameters for averaged and non-averaged BOLD signals as a function of the percent stimulus-related signal in the raw data. For extreme noise, where stimulus-related signal constitutes only 2 percent of the overall signal, averaging over 20 cycles (blue line) produces higher Similarity values than either non-averaged signal (black line) or signal averaged over 3 cycles (red line). For larger percentages of stimulus-related signal within the overall signal, averaging shows no effect.

Figure 2.5 shows the Similarity between the true underlying pRF parameters and those estimated from the simulated BOLD signals. For very noisy data, i.e. where signal resulting from stimulation constitutes merely 2 percent of the resulting BOLD signal (figure 2.3a), estimation from signal averaged over 18 cycles showed significantly larger Similarity values than estimation from both non-averaged signal and signal averaged over 3 cycles. For larger percentages of signal due to stimulation, however, the beneficial effect of extensive averaging disappeared. Following up on this result we investigated the effect of averaging in more detail for BOLD signals of which only 2 percent are due to averaging. To this end we varied the number of cycles over which averaging occurs in the range from 3 to 18 (in steps of 3). This analysis revealed that performance improved as averaging was performed for 3 to 9 cycles with mean Similarity values of $\mu = .69$ [.68 .70], $\mu = .72$ [.71 .73], and $\mu = .75$ [.74 .76] for 3, 6 and 9 cycles, respectively. Beyond 9 cycles averaging had no further beneficial effect with a mean Similarity value of $\mu = .75$ [.74 .76] for signals averaged over 18 cycles.

Consistency

In addition to stimulation procedure optimization we investigated the consistency of receptive fields across scanning sessions. To this end we correlated⁴ Polar Angle and Eccentricity maps obtained from different sessions and calculated the Similarity between pRF parameter estimates of the two sessions. We restricted our analysis to data obtained from presenting a bar stimulus in a random fashion interspersed with mean luminance periods. Additionally, we

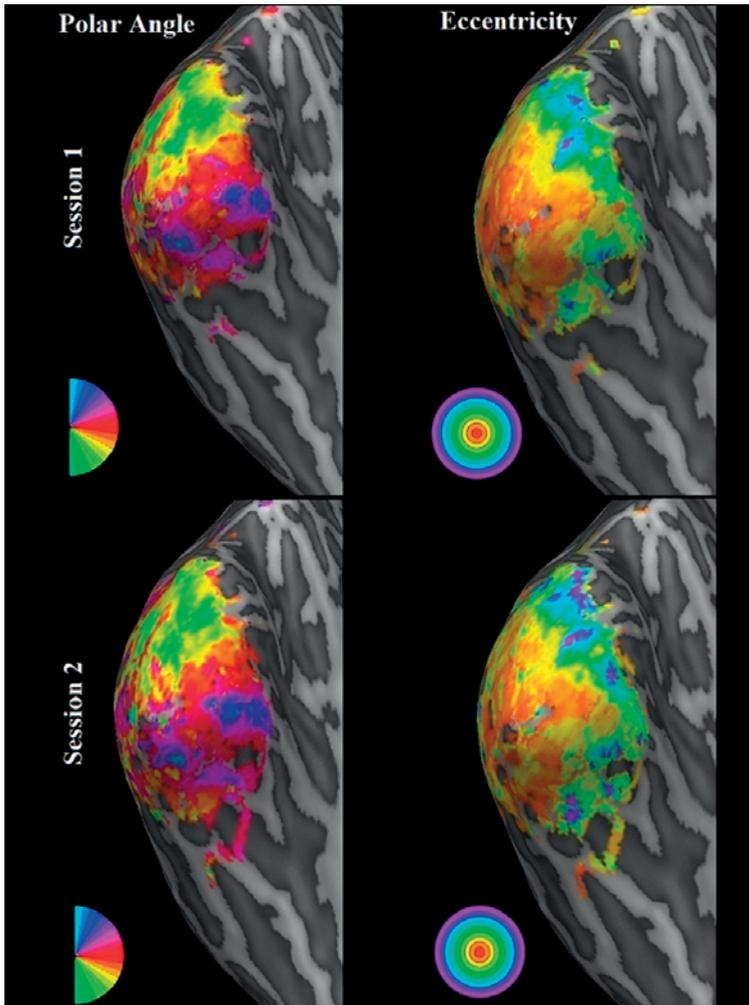


Figure 2.6: Retinotopy Surface Maps. This figure shows exemplary polar angle as well as eccentricity maps of subject 3. The maps were obtained from a randomly presented bar stimulus within two separate sessions. The upper row shows maps of the first session while the lower row shows maps of session two. In accordance with the correlation results between maps (see figure 2.8 for details), the two polar angle and eccentricity maps are highly visually similar.

⁴ We calculate circular correlation between Polar Angle Maps.

only considered voxels which had previously been selected for at least one of the two sessions. The means of all metrics were obtained by bootstrapping. To gain a clear picture which pRF parameters were more or less consistently estimated, we additionally used bootstrapping to obtain means of Similarity values resulting from successively leaving each parameter out. Figure 2.6 shows exemplary polar angle and eccentricity maps of subject 3 for the two ses-

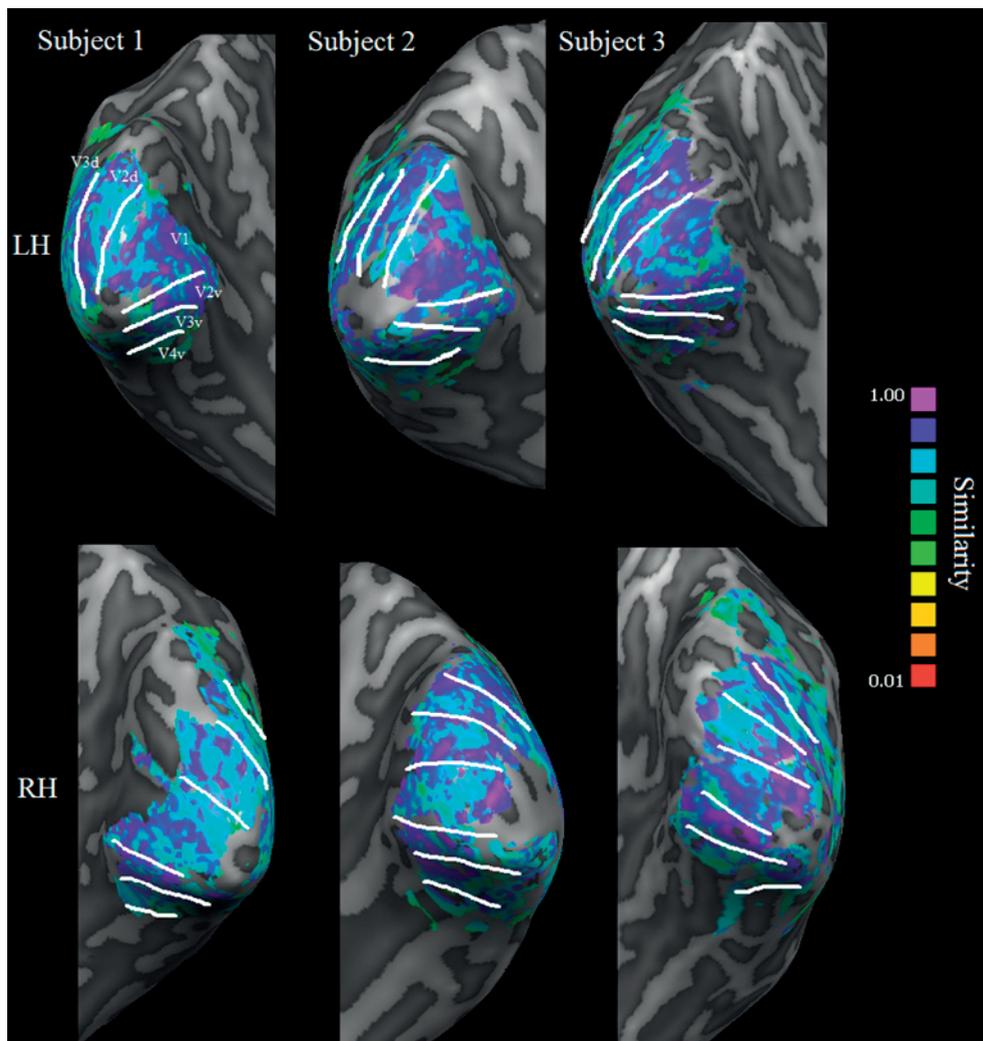


Figure 2.7: Similarity Surface Maps. This figure shows Similarity between pRF estimates obtained in sessions 1 and 2 for all three participants. The maps were obtained from a randomly presented bar stimulus within the two separate sessions. The upper row shows maps for the left hemisphere while the lower row shows maps for the right hemisphere. Borders delineating visual areas V1, V2, V3, and V4 obtained from the retinopy resulting from this stimulus are superimposed on the maps. The Similarity values show no systematic variation across subjects nor along the visual hierarchy.

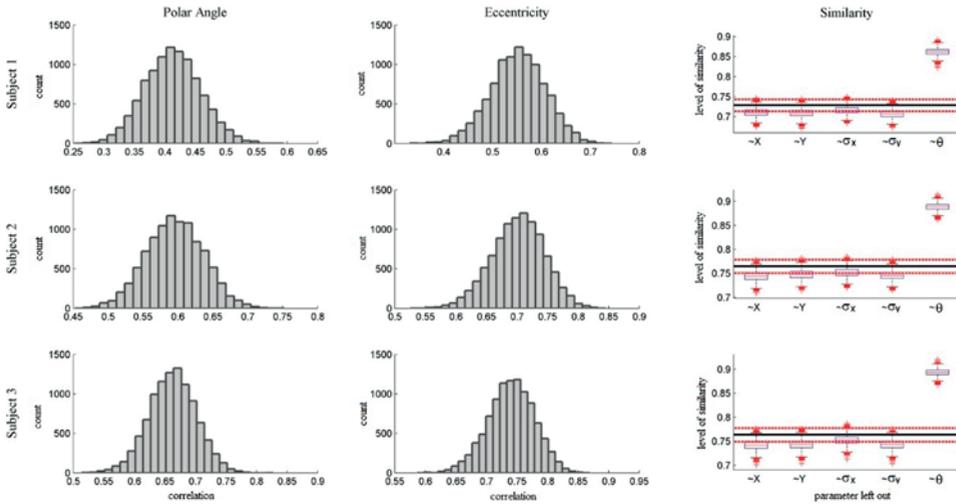


Figure 2.8: Temporal Consistency. This figure shows the correlation between polar angle and eccentricity maps obtained from two temporally separated sessions as well as the Similarity between pRF parameters estimated from those sessions. The results are shown individually for each subject (rows). For all subjects the distributions of correlations for both polar angle and eccentricity maps obtained using bootstrapping indicate a good fit between the maps. The final column shows Similarity values for all three subjects. The black horizontal line indicates the mean Similarity between the complete set of pRF parameters estimated from the two sessions. The red lines indicate confidence intervals around this mean as obtained via bootstrapping. The boxplots show the Similarity between a series of two sets of pRF parameters each time excluding one parameter in the Similarity calculation. While excluding any location or size parameter has no significant effect on Similarity, removing the angle of elongation significantly increases the Similarity between the two sets. This implies that location and size parameters are estimated very consistently. Angle of elongation, however, is subject to variability.

sions on an inflated cortex mesh while figure 2.7 shows surface maps detailing the Similarity values between pRF estimates of sessions 1 and 2. Results of the Similarity analyses are finally shown in figure 2.8.

All metrics indicated high consistency of pRF parameters across sessions. Specifically, polar angle maps showed a circular correlation of $r = .41$ 95% CI [.34 .49] for subject one, $r = .6$ 95% CI [.53 .66] for subject two, and $r = .66$ [.6 .73] for subject three. Eccentricity maps showed a correlation of $r = .55$ 95% CI [.46 .64] for subject one, $r = .7$ 95% CI [.63 .77] for subject two, and $r = .74$ 95% CI [.68 .8] for subject three. Finally, mean Similarity was $\mu = .73$ 95% CI [.71 .74], $\mu = .76$ 95% CI [.75 .78], and $\mu = .75$ 95% CI [.74 .77] for subjects one, two, and three, respectively. As can be seen from the third column of figure 2.8 removing any of the position or size pRF parameters in the calculation of the Similarity did not significantly affect the result. Leaving out the angle of elongation, however, caused the Similarity values to rise significantly. The mean Similarity of pRF location and size alone, i.e. leaving out angle of elongation, was $\mu = .86$ 95% CI [.85 .88], $\mu = .89$ 95% CI [.88 .9], and $\mu = .89$ 95% CI [.88 .9] for the three subjects, respectively. This implies that while location and size parameter estimates were highly consistent across sessions, the angle of elongation estimate changed somewhat across sessions. To establish

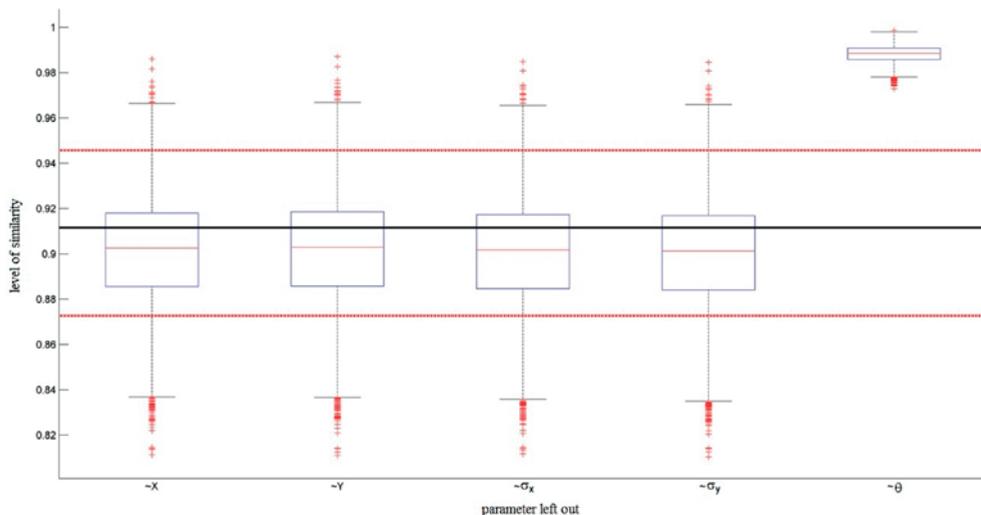


Figure 2.9: Similarity between known and estimated pRF parameters. This figure shows the Similarity between the entire sets of known and estimated pRF parameters (horizontal lines, black = mean, red = confidence interval obtained via bootstrapping) and the Similarity between a series of sets while each time excluding one parameter (boxplots). While excluding any location or size parameter has no significant effect on Similarity, removing the angle of elongation significantly increases the Similarity between the two sets. This implies that location and size parameters can be estimated very accurately. The angle of elongation parameter, however, diverges somewhat from ground truth.

whether this was due to dynamic changes in the true underlying receptive fields or simply due to a general difficulty in estimating this parameter we performed bootstrapping on the Similarity values between true pRF parameters and those estimated from simulated data using the same stimulus. The results are shown in figure 2.9. These results showed the same overall pattern with highly accurate estimations of location and size parameters but not of the angle of elongation. This implies that the inconsistencies across sessions with regard to that parameter were likely caused by a general difficulty in estimating it rather than underlying dynamic changes. This result is in line with previous findings of Lee *et al.* (2013) who showed that directly fitting an anisotropic pRF model to fMRI data cannot identify the angle of elongation with an equal degree of precision as for the other parameters.

DISCUSSION

The present study pursued two goals. One of these goals concerned the identification of best practices with regard to the estimation framework in the context of population receptive field mapping. The other concerned the validity of the assumption that receptive fields are static or temporally consistent. With regard to the former goal our results show that a

bar stimulus yields a set of pRF parameters from which BOLD activity resulting from different visual stimulation protocols can be reliably predicted. Additionally, simulation results, for which all assumptions regarding the underlying pRF structure are met by design, show that pRF estimates obtained from stimulation using a bar are more accurate than the estimates obtained from stimulation using a wedge-ring. For these reasons we recommend a bar stimulus. Additionally, the merit of pRF estimates increases for any stimulus, if it is presented in a random fashion. Furthermore, pRF estimation in areas with large receptive fields strongly benefits if stimulation is interspersed with mean luminance periods while estimation of early visual cortex is hardly affected. We, therefore, recommend the use of a bar stimulus presented randomly and interspersed with mean luminance periods for pRF mapping procedures.

With regard to data preparation the benefit of averaging the BOLD signal across stimulus cycles depends on the level of noise in the data. If noise is low, averaging conveys no advantage in terms of the precision with which pRF parameters are estimated even though a larger proportion of variance in the BOLD signal is explained. Only for extremely noisy data; i.e. for very small voxels, can averaging provide more accurate pRF parameter estimates. In order for averaging to take effect, however, a sufficiently large number (~ 9) of stimulus cycles is needed. Nonetheless, averaging is accompanied by benefits other than precision in estimation. First of all, the increasing amount of variance that can be explained for averaged as compared to non-averaged signals provides a straightforward demarcation of those voxels processing visual information from those that do not. Another benefit is that averaging the signal reduces the amount of data points and hence reduces the time needed for the estimation process. For these reasons we generally recommend signal averaging and regard it essential for high resolution fMRI data. This requires that the randomization is identical for all stimulus cycles thus limiting the number of possible cycles before the order becomes predictable. An alternative approach might be to concatenate stimulus cycles rather than to average across them. This would allow full randomization and maintain an unpredictable order of stimulus presentations. How such an approach compares to straightforward averaging is an interesting question which remains to be elucidated.

In order to further improve the estimation framework, it is possible to extent stimulation to incorporate different types of aperture content. Usually, stimulus aperture reveals a flickering checkerboard with vertical and horizontal borders. It might be interesting to vary the orientation of these borders in order to accommodate for the radial bias hypothesis (Sasaki et al., 2006) and to examine how this affects pRF parameter estimates. Additionally, for the study of pRFs of late visual areas which react primarily to complex visual stimuli it might prove fruitful for the stimulus aperture to reveal complex shapes, motion, or contours (Dumoulin et al., 2014). Furthermore, or in order to tailor the pRF estimation framework to more specific research questions such as disease, it might be necessary to further adapt the stimuli. It had previously been shown, for instance, that randomly presented multifocal stimuli improve pRF estimation in the presence of foveal scotomas (Binda et al., 2013).

Our results with regard to the second goal show that estimated pRF parameters appear to be generally very consistent over time. This is especially true for pRF location and size for which fluctuations in parameter values across sessions are within a very narrow range. With regard to the angle of elongation of anisotropic receptive fields fluctuations in parameter values appear to fluctuate within a broader range. The consistency of angle of elongation is thus somewhat lower than that of other parameters. A possible remedy for this is to employ a pRF mapping approach without a priori assumptions with regard to the pRF shape, as proposed by Lee *et al.* (2013). This might narrow the range of acceptable values for angle of elongation and consequently boost this parameter's temporal consistency. Another important methodological consideration is that Similarity values observed empirically are lower than those observed for simulations. This is attributable to the fact that the parameter space was exhaustive for simulations by design while it can hardly be considered to cover the entire range of possible parameter values among voxels. It is thus conceivable that where voxels have an actual receptive field somewhat between two or more points in the examined parameter space it might be attributed to one of them stochastically and thus differ from session to session.

Overall our results suggest that a randomly presented bar stimulus interspersed with mean luminance periods should be used for the estimation of population receptive fields. Ideally but not necessarily, the resulting data should be averaged. Receptive field parameters thus obtained are temporarily stable and can be utilized for numerous applications without hesitation.

ACKNOWLEDGEMENTS

This project was supported by the European Research Council under the European Union's Seventh Framework Programme (ERC-2010-AdG, ERC grant agreement no. 269853).

Chapter

3

Comparing Computational Models of Simultaneous Brightness Contrast Using Ultra-High Field Functional Magnetic Resonance Imaging

This chapter is an adapted version of the following article:

Senden, M., Peters, J., Jans, B., Reithler, J., & Goebel, R. (in preparation). **Comparing Computational Models of Simultaneous Brightness Contrast Using Ultra-High Field Functional Magnetic Resonance Imaging.**

ABSTRACT

Functional magnetic resonance imaging (fMRI) has in recent years become a viable tool for comparative theory testing. A growing number of studies compare information classification results from (outputs of) various simulations to observed fMRI activity, to gain insights in the type of information represented in dedicated brain regions. However, to gain a deeper understanding of neural mechanisms it is often necessary to go beyond evaluating which information is represented throughout various processing stages and to focus on how it is represented. In the present study we demonstrate how empirical and simulated fMRI data can be integrated in the same reference space to compare the detailed spatial activation profile of representations predicted by competing computational models. We build on the previously proposed common brain space (CBS) which is an idealized, retinotopically organized, surface representation of visual cortex. We extend the CBS framework by incorporating population receptive field estimates as a means to establish functional correspondence between voxels, model units, and CBS locations at a resolution sufficient to contrast spatial activation profiles of different computational models. We exemplify the approach by comparing two “brightness induction” models accounting for a simultaneous brightness contrast effect. The first model accounts for the effect in terms of spatial filtering in conjunction with contrast normalization. The second model proposes an edge integration mechanism wherein representations of surface luminance are modulated by lateral spread of edge information. We find that the outcome of the model comparison depends on the direction of brightness induction suggesting that the brain might utilize aspects of both mechanisms. In the future, our approach might aid in developing a model reflecting a hybrid mechanism by evaluating candidate implementations against each other in terms of the spatial activation profiles they imply.

INTRODUCTION

It has been repeatedly suggested that comparative theory testing is a commendable practice in conducting research (Chamberlin, 1890; Clarke, 2007; Lakatos, 1976; Platt, 1964). Not merely because adopting several theories might alleviate the threat of “parental affections” (Chamberlin, 1890) towards a specific theory and of the resulting confirmation bias, but also because it is crucial for resolving theoretical controversies (Platt, 1964). However, it is usually not sufficient to contrast theories within a single methodological domain since crucial findings often come with a transition to novel methods (Chatham & Badre, 2014; Greenwald, 2012). Recent neuroimaging studies have, for instance, greatly contributed to the resolution of long-standing theoretical controversies in psychology (Greenwald, 2012). It is thus unsurprising that functional magnetic resonance imaging (fMRI) is becoming an increasingly important tool for comparative theory testing (Chatham & Badre, 2014; Greenwald, 2012; Mather, Cacioppo, & Kanwisher, 2013). One fruitful approach is to use fMRI-based decoding techniques to contrast theories and models which either make different predictions with respect to the information represented across a set of brain regions (e.g. O’Craven, Downing, & Kanwisher, 1999) or with respect to the information represented within a brain region (e.g. Khaligh-Razavi & Kriegeskorte, 2014).

While the decoding approach is vital for comparative theory testing, it cannot accommodate cases wherein competing models make discriminative predictions not with respect to what is represented but rather how it is represented; i.e. with respect to the detailed spatial activation profile of representations within a brain region. The previously proposed common brain space (Johnston, Boehm, Healy, Goebel, & Linden, 2010; Peters et al., 2010, 2012) provides a framework suited for such cases in early visual cortex. The CBS is an idealized, retinotopically organized, surface representation of visual cortex which allows for the assignment of empirical but also simulated data based on anatomical or functional correspondence between voxels, model units, and CBS locations. Previously this assignment was based on localizer analyses providing coarse retinotopic information (Peters et al., 2010). To improve spatial detail of the CBS approach we utilize the recently developed population receptive field (pRF) mapping procedure (Dumoulin & Wandell, 2008) and a complex-logarithm transformation of visual field to cortical coordinates (Balasubramanian, Polimeni, & Schwartz, 2002; Schwartz, 1980). Since cortical coordinates of V1 voxels determine their receptive field properties (Holmes, 1918; Sperry, 1963), they can be functionally inferred to a large degree of precision based on pRF coordinates measured at high spatial resolution ($0.8 \times 0.8 \times 0.8 \text{ mm}^3$) using ultra-high-field fMRI. The resulting functional alignment allows for the straightforward integration of data across subjects. At the same time, definition of the CBS in cortical coordinates provides a direct quasi-anatomical link for neural network models. In this way, the CBS provides a reference frame for the comparison of detailed spatial activation profiles predicted by competing models.

We exemplify this framework in the context of simultaneous brightness contrast, a visual illusion wherein an equiluminant gray patch appears darker when presented on a white background than on a black background (Chevreul, 1839; Hess & Pretori, 1894). Neurophysiological as well as fMRI research has established that this illusion, and brightness perception in general, is reflected in signals observed in early visual cortex (Kinoshita & Komatsu, 2001; MacEvoy, Kim, & Paradiso, 1998; Meng, Remus, & Tong, 2005; Rossi, Rittenhouse, & Paradiso, 1996; van de Ven, Jans, Goebel, & De Weerd, 2011). However, the mechanism underlying the effect remains contested (Kingdom, 2011). One prominent class of models proposes spatial filtering as the fundamental underpinnings of surface perception (Blakeslee, Cope, & McCourt, 2015; Blakeslee & McCourt, 1999, 2004; Dakin & Bex, 2003) while another proposes an edge integration mechanism (Hurlbert & Poggio, 1988; Rudd & Arrington, 2001; Rudd & Popa, 2007; Rudd, 2001). Here we implement two neural network models each expressing the basic principles of one of these classes. With respect to the spatial filtering approach we implement a neurally extended version of the Oriented Difference-of-Gaussian (ODOG) model developed by Blakeslee, McCourt, and colleagues (Blakeslee et al., 2015; Blakeslee & McCourt, 1999, 2004). Briefly, surface perception results from the output of a set of orientation and spatial frequency tuned linear filters (Blakeslee et al., 2015; Blakeslee & McCourt, 1999, 2004). Illusions with respect to brightness occur due to the attenuation of low spatial frequencies as well as from a contrast normalization step. With respect to the edge integration approach we consider a model recently developed in our laboratory (Peters et al., 2010). Briefly, the response of units reflecting luminance levels in the visual input are modulated by lateral excitation and inhibition spreading from edge-detection units.

While both models can account for the perceptual effect of simultaneous brightness contrast, they make differential predictions for the detailed spatial profile of fMRI activity. Here we use the common brain space approach to integrate high-resolution fMRI data of three subjects in order to compare the two models. Specifically, after replicating previous findings of our group showing the existence of an fMRI signal related to the simultaneous brightness contrast illusion in early visual cortex (van de Ven et al., 2011), we project fMRI data to the common brain space. Subsequently we use one of two sets of ultra-high-field fMRI data collected in a simultaneous brightness contrast experiment to adjust parameter values of the two models. Finally, we compare the fitted models against the second dataset by evaluating each model's relative quality using the Akaike information criterion (Akaike, 1973).

MATERIALS AND METHODS

Subjects

Three healthy subjects (1 female, age range = 30-38 years, mean age = 34.67 years) participated in this study. All subjects had normal or corrected-to-normal visual acuity and provided

written informed consent according to a protocol approved by the local research ethics committee (local Ethical Committee of the Faculty of Psychology and Neuroscience at Maastricht University).

Stimuli and Task

Subjects were instructed to maintain fixation throughout the experiment, which comprised a psychophysical rating task prior to scanning, one control and two experimental runs of simultaneous brightness contrast, one stimulus localizer run, one pRF mapping run, and one run used for estimating hemodynamic responses.

Simultaneous Brightness Contrast. In accordance with previous fMRI studies, we used a dynamic version of the simultaneous brightness contrast stimulus (De Valois, Webster, De Valois, & Lingelbach, 1986; van de Ven et al., 2011). An annulus (the “probe” stimulus) of constant luminance was interspersed between a central disk and a peripheral annulus (the “inducers”) whose luminance changed in a quasi-sinusoidal fashion from white (up) to black (down; see figure 3.1). The probing region was gray (23.38 cd/m²) and black (.75 cd/m²) for an experimental condition with and a control condition without brightness induction, respectively. The two inducers dynamically varied across the entire luminance range (from .75 cd/m² to 48.28 cd/m²). The radius of the inner inducer was .46°, the width of the probing region was 1.99° (radius at outer edge = 2.45°), and the width of the outer inducer was 5.56° (radius at outer edge = 8°). These values were obtained from an inverse of the topographic mapping of the visual field onto the cortical surface (Balasubramanian et al., 2002; Schwartz, 1980) using average parameter values ($a = .7^\circ$, $\alpha = .9$) reported by Polimeni *et al.* (2005). The corresponding values in cortex coordinates were: a radius of 7.5 mm for the inner inducer and a width of 15 mm for both the probe and outer inducer (radii at the outer edges were equal to 22.5 mm and 37.5 mm, respectively). Stimuli were presented concentrically around a fixation cross of .37° width and height.

In each experimental run, 278 functional volumes were acquired (~11 min; duration of each volume = 2106ms) consisting of 270 volumes of dynamic luminance changes in the inducer regions preceded and followed by a rest period of 4 volumes. Inducers changed luminance quasi-sinusoidally at .083 Hz with 6 rest volumes (gray screen) interspersed after each half period. This stimulation pattern results in 15 repetitions of both the gray to white (inducer-up) and gray to black (inducer-down) luminance brightness changes (see figure 3.1). We acquired data for two experimental and one control run.

Stimulus Localizer. Three circular checkerboard stimuli of equal radii and widths as the inner inducer, probe, and outer inducer stimuli were used. Checkerboards reversed contrast at a rate of 10 Hz. The localizer run comprised 108 volumes showing six repetitions of each of the three stimulus locations for four volumes followed by a resting period of two volumes.

Additional rest periods of two volumes were shown at the beginning and end of the run leading to a total of 112 volumes (~4.5 min).

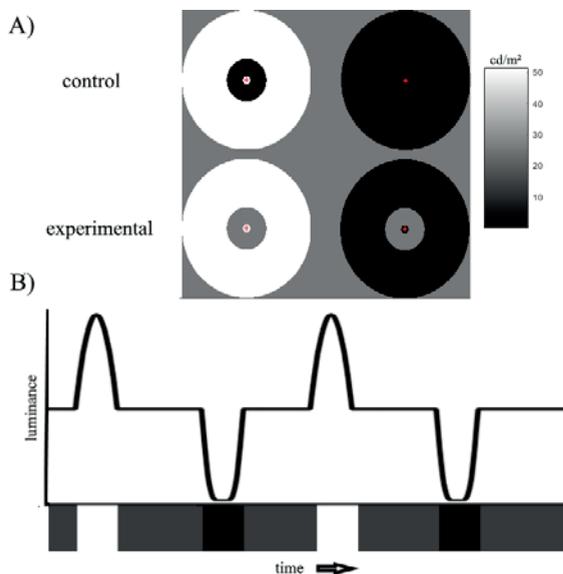


Figure 3.1: Stimulation. *Panel A)* shows the ring stimuli for the control and experimental conditions used in the simultaneous brightness contrast fMRI runs in terms of physical luminance. *Panel B)* shows an extract of the first 90 seconds of the stimulation procedure. Gray regions reflect presentation of an isoluminant screen, white regions a luminance change in the inducers from gray to white and back, and black a change from gray to black and back. The curve shows the exact shape of luminance changes. Luminance change was coded in RGB color values. Given the nonlinear relationship between RGB color values and true luminance, up and down sweeps are of slightly different shape. We took these differences into account in our analyses.

pRF mapping. A bar aperture (1.33° wide) revealing a checkerboard pattern was presented in four orientations. For each orientation the bar covered the entire screen in 12 discrete steps (one per volume). Within each orientation the sequence of steps (and hence of the locations) was randomized (Senden, Reithler, Gijssen, & Goebel, 2014). Each orientation was presented six times leading to a total of 288 volumes of stimulus presentation preceded and followed by a rest period of eight volumes leading to a total of 304 volumes (~12 min).

Estimation of hemodynamic response function. A full-field checkerboard reversing contrast at a rate of 10 Hz was repeatedly (10 times) presented for a single volume followed by a 10 volume resting period. The resulting 110 volume protocol was preceded and followed by a two volume rest period (114 volumes; ~4.5 min).

Stimulus Presentation

The bar stimulus used for pRF mapping was created using the open source stimulus presentation tool BrainStim (<http://svengijssen.github.io/BrainStim/>) and presented at a resolution of 1200x1200 pixels (screen resolution was 1600x1200 pixels). The remaining stimuli were created in Presentation® (Version 15.0; www.neurobs.com) and presented at a resolution of 1024x768 pixels. Stimuli were projected on a frosted screen at the top end of the scanner bed using an LCD projector (Panasonic, No PT-EZ57OEL; Newark, NJ, USA) and reflected by a mirror mounted to the head coil.

Psychophysical Ratings of Induced Brightness

Subjects rated the strength and direction of induced brightness change in an offline psychophysical experiment while being situated in the scanner bore. Subjects adjusted the physical luminance value in the probe region of a test stimulus to correspond to the perceived luminance experienced for simultaneous brightness contrast conditions. In each trial, the inducer region changed from gray to white (black) within three seconds. Upon reaching its maximum (minimum) luminance, the inducer but not the probe region was masked by a flickering checkerboard. Next, subjects used two buttons to adjust the luminance of the probe region to their percept experienced during the inducer-up (-down) conditions. At the start of the following trial, the probe region was reset to gray. Subjects completed 30 trials for the inducer-up and inducer-down conditions each. The resulting ratings were normalized with respect to the full range of luminance values.

Magnetic resonance imaging

Images were acquired with a Siemens MAGNETOM 7 T scanner (Siemens; Erlangen, Germany) and a 32-channel head-coil (Nova Medical Inc.; Wilmington, MA, USA). In-session anatomical data were collected with a T1-weighted magnetization prepared 2 rapid gradient echoes (MP2RAGE; Marques et al., 2010) sequence (240 sagittal slices, TR = 5000 ms; Echo Time (TE) = 2.47ms; Inversion Time (TI1/TI2) = 900/2750 ms; Flip angle (FA1/FA2) = 5°/3°; generalized autocalibrating partially parallel acquisitions [GRAPPA] acceleration factor = 2; Field of View [FoV] = 224 × 203 mm², voxel size = .7 × .7 × .7 mm³). In addition, we acquired a high resolution anatomical scan at .4x.4x.8mm³ covering the occipital lobe which was used for segmentation of the white-gray matter boundary (MP2RAGE; 88 coronal slices, TR = 5000 ms; TE = 2.54ms; TI1/TI2 = 900/2750 ms; FA1/FA2 = 5°/3°; FoV = 180 × 162 mm²).

Functional measurements comprised one control and two experimental runs of simultaneous brightness contrast, one eccentricity localizer run, one pRF mapping run, and one run used for estimating hemodynamic responses. To correct for EPI distortions, five additional functional volumes with reversed phase encoding direction were acquired at the start of each session (Andersson, Skare, & Ashburner, 2003). High-resolution functional images were obtained using gradient echo (T2* weighted) 3D echo-planar imaging (3D-EPI; Poser, Koop-

mans, Witzel, Wald, & Barth, 2010; 44 oblique coronal slices; TR = 2106 ms; TE = 18 ms; FA = 14°; GRAPPA acceleration factor = 3; FoV = 148 x 148 mm²; voxel size = .8 x .8 x .8 mm³).

Processing of (f)MRI data

All imaging data were analyzed using BrainVoyager QX (v2.8; Brain Innovation, Maastricht, the Netherlands) as well as custom code developed in MATLAB (2015a, The MathWorks, Natick, MA). Anatomical images were skull-stripped and corrected for intensity inhomogeneities by estimating a bias field based on a preliminary white matter segmentation (Vaughan et al., 2001). The resulting dataset was subsequently interpolated to a nominal voxel size of .4 mm isotropic to match a multiple of two of the resolution of the functional data. We furthermore aligned high resolution anatomical images (interpolated to .4 mm isotropic) acquired in previous sessions to the newly acquired anatomical images. Aligned, high resolution anatomical images were segmented according to boundaries between gray and white matter after which the cortical sheet of each hemisphere was tessellated into a 3D cortical surface representation (Goebel, Esposito, & Formisano, 2006; Kriegeskorte & Goebel, 2001).

Functional runs were preprocessed using standard steps including 3D rigid body motion correction (aligning all runs to the first functional run) and temporal high-pass filtering using a general linear model (GLM) Fourier basis set of two cycles sine/cosine per run (including linear trend removal). Subsequently, 5 additional functional volumes of reversed phase encoding were used to correct distortions in echo-planar images. In these images distortions occur in the opposite direction from those of our functional runs and can be used to estimate the susceptibility-induced off-resonance field using a method similar to that described in (Andersson & Skare, 2002) as implemented in the COPE v.03 plugin for BrainVoyager QX. The resulting off-resonance field estimation was used to correct for EPI distortions. Finally, functional runs were co-registered to individual anatomical scans and volumetric timecourses were resampled to the cortical surface and z-normalized.

Analyses of fMRI Data

pRF Mapping and V1 Delineation. For each subject, we mapped population receptive fields using the standard pRF model described by Dumoulin and Wandell (2008) and performing a grid search to obtain optimal parameter values. In terms of pRF location, the visual field was split into a grid of 101 by 101 points covering 16° of visual angle in both the X and Y directions. Since we were mainly interested in pRF locations of V1 vertices, we restricted the set of pRF sizes to 9 values ranging from .2° to 2° of visual angle. The pRF coordinates resulting from the estimation step were converted to polar angle values as well as to cortical coordinates using the complex-logarithm transformation described by (Balasubramanian et al., 2002; Schwartz, 1980) and parameter values given in (Polimeni et al., 2005). Next, polar angle maps presented on inflated cortical renderings were used to delineate a V1 region-of-interest in each hemisphere. Vertices were selected for further analysis if they fell within either left or right V1 and if

at least 30% of variance in their pRF timecourse was explained. This resulted in 17837, 15367, and 18486 vertices for the three subjects, respectively. These numbers represent 83.5%, 74.4%, and 75.9% of vertices falling within the V1 region of interest for the three subjects before the additional requirement of explaining at least 30% of variance was applied.

Hemodynamic Response Estimation. We estimated the subject- and vertex-specific hemodynamic responses by performing a regression analysis with a set of 11 finite impulse responses (FIR) as a predictor of the BOLD signal observed for each vertex in the hemodynamic response estimation run. In the remaining analyses, predictors were individually convolved with their specific hemodynamic response.

Stimulus Localization. Three predictors modeling the sequence of blocks of stimulus positions (effect coding) were used to identify the inner inducer, probe, and outer inducer regions. Results of the full regression model were FDR-corrected ($q = .05$) for multiple comparisons. Vertices failing to reach significance were discarded in further analyses. However, since the analysis was restricted to vertices selected based on the V1 delineation step described earlier, 95%, 99%, and 91% of vertices reached significance and could thus be retained for the three subjects, respectively. Following linear regression we performed contrast analyses to determine the regions where each of the three circular stimuli evoked stronger responses than the other two stimuli.

Simultaneous Brightness Contrast. In the simultaneous brightness contrast experiment, we performed a fixed-effects group regression analysis in separate regions of interest (ROIs). We defined two different types of ROIs: 1) the inducers and probe regions as defined by the stimulus localization procedure described above (i.e., 3 "Stimulus ROIs"), 2) 30 bands of equal width representing increasing cortical distance defined using functionally derived cortical vertex coordinates (i.e., 30 "Cortical distance ROIs"). The analysis comprised three separate predictors. The first predictor reflected luminance values relative to gray presented in the inducer regions over time (dummy coding, $0 = baseline$). For each vertex, this predictor was convolved with its hemodynamic response. The second predictor was categorical and indicated control vs first experimental fMRI runs (dummy coding, $0 = control$). The third predictor reflected the interaction between predictors one and two. We assume that a simultaneous brightness contrast effect in the probe region would be reflected by this predictor since prior research has shown that surface luminance loses its predictive power in the experimental as compared to control condition (van de Ven et al., 2011). Furthermore, we also included a covariate reflecting the presence of luminance change irrespective of the direction of this change (dummy coding, $0 = baseline$). Similar to the predictors, the covariate was convolved with each vertex's hemodynamic response.

Computational Modeling

Oriented Difference-of-Gaussian Model (ODOG). This model creates a brightness percept in three stages of processing (Blakeslee et al., 2015; Blakeslee & McCourt, 1999, 2004). In the first stage, visual input (presented as a 256x512 pixel grid spanning 16° visual angle horizontally and 32° vertically) is convolved with a set of linear filters which are orientation and spatial frequency selective in accordance with oriented difference-of-Gaussians. These filters might be considered to be the receptive fields of retinotopically organized simple cells in a single hemisphere (DeValois, DeValois, & others, 1988; Hubel & Wiesel, 1959; Movshon, Thompson, & Tolhurst, 1978b). In total six different orientations (0°, 30°, 60°, 90°, 120°, and 150°) are paired with seven spatial frequencies. Spatial frequencies are arranged at octave intervals (specific values in our model implementation were controlled by a free parameter; see table 2). In the second stage, the outputs of spatial frequency filters are integrated within each location and orientation with a weighting attenuating low spatial frequencies (Shapiro, Knight, & Lu, 2010). This integration might be considered to occur in complex cells (Movshon, Thompson, & Tolhurst, 1978a). In the third stage, the integrated response for each orientation is normalized by its root mean square magnitude akin to a contrast normalization process (Bonin, Mante, & Carandini, 2005). The resulting activation was fed into a logistic activation function to generate the final unit output reflecting the percept as well as the spiking activity observed at each location in the model. Since our aim was to compare model to fMRI activations which to a large extent reflect (subthreshold) synaptic rather than spiking activity (Logothetis et al., 2001; Maier et al., 2008), it was essential to estimate synaptic activity as well. We considered this to be reflected by the absolute of the total excitatory and inhibitory input which in this case amounts to the average of the weighted absolute filter responses across all orientations and spatial frequencies multiplied (rather than divided) by the root mean square magnitude. The exact contribution of spiking and synaptic activity to the final fMRI-like activity was controlled by a free parameter. We restricted our simulations to steady state responses to a gray probe surrounded by white inducers as well as a gray probe surrounded by black inducers. For more details on the implementation of the general ODOG model see Blakeslee (Blakeslee et al., 2015).

We have made three important changes to improve neural validity of the model. Firstly, V1 cortical magnification is taken into account such that the model has a larger proportion of units with receptive fields located in the foveal as compared to the peripheral regions (Balasubramanian et al., 2002; Polimeni et al., 2005; Schwartz, 1980). Secondly, spatial frequency selectivity is expressed in cycles per millimeter rather than cycles per degree. This is done to ensure that filters remain identical across the modeled cortical sheet in order to facilitate the filter process. At the same time, in conjunction with cortical magnification this accounts for units being tuned to progressively lower spatial frequencies with increasing eccentricity and is thus biologically plausible (Henriksson, Nurminen, Hyvärinen, & Vanni, 2008; H.-H. Yu et

al., 2010). Thirdly, following Robinson *et al.* (2007), local normalization is performed for each orientation; i.e. restricted to a region described by a Gaussian centered on each unit.

The model had five free parameters which were explored across a range of plausible values (see table 2) and selected based on how well the resulting model fits a validation data set of the simultaneous brightness contrast experiment. A parameter λ controlled the lowest of the seven spatial frequencies (the remaining were the next six octaves). Two parameters, a factor g and an exponent α , controlled the weighting of spatial frequencies. A parameter σ_0 controlled the size of the Gaussian region for local normalization. Finally, a parameter W_0 controlled the relative contribution of spiking (W_0) and synaptic ($1-W_0$) activity to the overall steady-state response.

Edge Detection Model (EDGE). This model gives rise to a brightness percept via the interplay of three retinotopically organized processing layers. As for the ODOG model, V1 cortical magnification is taken into account in the arrangement of V1 units. All layers represent V1 in a single hemisphere. The first layer represents a model retina, whereas layers two and three represent blob- and interblob-like units in V1 (Grossberg, 2003; Peters *et al.*, 2010). The retinal layer is represented by a 512x1024 pixel grid (spanning 16° visual angle horizontally and 32° vertically) and signals physical surface luminance. The V1 interblob layer is represented by a 256x512 pixel grid spanning the same visual angle as the retinal layer. Units in this layer are edge selective (Hubel & Wiesel, 1959; Yoshioka, Blasdel, Levitt, & Lund, 1996); i.e. they respond to a specific combination of orientation and contrast polarity. Specifically, receptive fields of these units sample four units from the retinal layer with two excitatory inputs on the left, right, bottom, or upper part of the receptive field corresponding to left, right, bottom, and top edge selective units, respectively. The V1 blob layer is, like the V1 interblob layer, represented by a 256x512 pixel grid. Units in this layer integrate the luminance information from the retinal layer with edge information from the V1 interblob layer. Each unit in the V1 blob layer receives excitatory input from four units of the retinal layer as well as excitatory and/or inhibitory input from V1 interblob units. Inspired by findings that lateral connectivity mediates context modulation (i.e. facilitation/suppression; Jones, Grieve, Wang, & Sillito, 2001; Li & Li, 1993) and that suppression can be spatially asymmetric (Walker, Ohzawa, & Freeman, 1999), edge detector units have net excitatory lateral connections to V1 blob units located in the darker region demarcated by the edge and net inhibitory lateral connections to V1 blob units located in the brighter region demarcated by the edge. The extent of lateral connectivity decays according to a Gaussian distribution centered on the edge detection unit. The original EDGE model also included non-specific lateral inhibition within the V1 blob layer. However, since this had no discernible effect on simultaneous brightness contrast, we chose to omit it. As for the ODOG model, we derived both spiking and synaptic activity from the model to generate the final fMRI-like activity. The spiking activity (output) of each unit is given by applying a logistic function to its inputs. Synaptic activity of each unit is considered to be the absolute of its

total excitatory and inhibitory input. We restricted our simulations to steady state responses to a gray probe surrounded by white inducers as well as a gray probe surrounded by black inducers. For more details on the implementation of the EDGE model see (Peters et al., 2010).

The model had five free parameters whose values were explored across a range of plausible values (see table 2) and selected based on how well the resulting model fits a validation data set of the simultaneous brightness contrast experiment. A parameter σ_E controlled the width of the Gaussian reflecting the extent of lateral impact of V1 interblob on V1 blob units. Three parameters, b_1 , b_2 , and b_3 , controlled the threshold in the logistic activation function of the retinal, V1 interblob, and V1 blob units, respectively. Finally, a parameter W_E controlled the relative contribution of spiking (W_E) and synaptic ($1-W_E$) activity to the overall steady-state response.

Common Brain Space. In order to have a reference frame for integrating empirical data of several subjects as well as simulated data, we defined a common brain space of V1. The CBS is a cortical coordinate system whose X-axis follows the representation of the horizontal meridian along the calcarine sulcus and whose Y-axis follows the polar angle gradient from the horizontal to the upper and lower vertical meridians (see figure 3.2a; Balasubramanian et al., 2002; Schwartz, 1980). Within this CBS representation, brain hemispheres were abstracted away such that two points with identical coordinates relative to the horizontal and vertical meridians in their respective hemispheres, were collapsed onto a single point in the common brain space. The X-axis ranged from zero (foveal representation) to 40 mm in 31 steps; the Y-axis ranged from minus (dorsal) to plus (ventral) 18 mm in 31 steps. We integrated data of all subjects into the common brain space in two steps. First, within a subject, we calculated Euclidean distances between CBS coordinates and each vertex's presumed cortical coordinate given its pRF properties and assigned vertices to locations in the CBS if their distance was below 2 mm. Figure 3.2b shows the average percentage of vertices assigned to each CBS location, the maximum being five percent. Common brain space locations not being assigned any vertices were neglected during all analyses. Each vertex was assigned to between 4 and 10 (of 961) CBS locations. The contribution of vertices thusly selected was further weighted according to a Gaussian function of the Euclidean Distances with a spread of $\sigma = 1$ mm. Timecourses at each CBS coordinate were the weighted average of timecourses observed for connected vertices. Second, the resulting CBS data were averaged across subjects. Next, we obtained empirical spatial activation profiles for inducer-up and inducer-down conditions at each CBS location by computing the dot product of each condition's event related average observed at that location and an idealized event related response given by a grand-average hemodynamic response (average over all subjects and vertices) and the stimulus on- and off-sets (see figure 3.3 for details). This dot product essentially produces a weighted average with the largest weighting for points close to the expected peak response. For simulated data,

we also followed a Euclidean distance-based approach in assigning model units to points in the CBS.

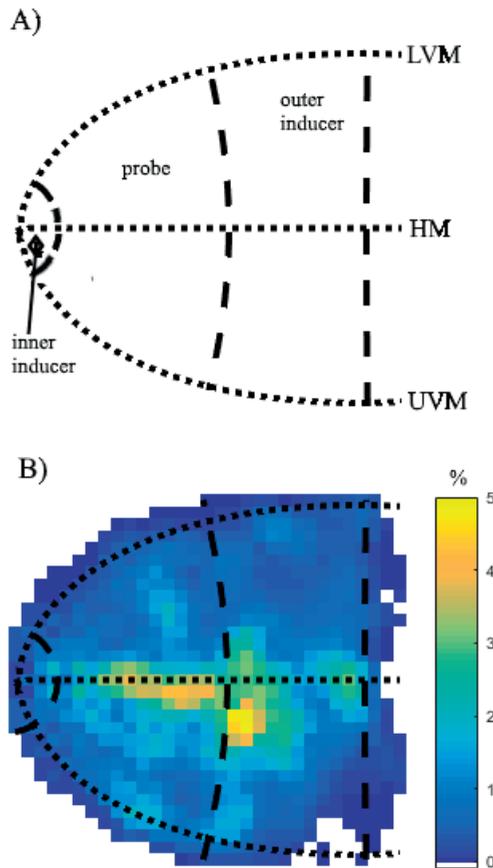


Figure 3.2: Common brain space. *Panel A)* shows a schematic representation of the common brain space for V1 with outer borders of the three rings superimposed. Dashed lines indicate, from left to right, the outer borders of the inner inducer, probe, and outer inducer stimulus regions. The dotted lines reflect where the horizontal (HM) as well as upper (UVM) and lower (LVM) vertical meridians are represented in the CBS. *Panel B)* shows vertex sampling in terms of the average percent of V1 vertices in a subject assigned to each location in the common brain space. The outer edges of the CBS were not assigned any vertices (white region). In general all parts of the CBS sample similar amounts of vertices (between 1 and 5 percent). However, the largest number of vertices contribute to CBS locations along the horizontal meridian (~5%).

However, since the resolution of the model cortical sheet is higher, especially for the foveal region, only the unit with the lowest distance was assigned to each CBS location.

Model Parameter Selection. For both the ODOG and the EDGE model five parameters were fit such that the model predictions produced a good correspondence with a validation data

set (experimental run 1). Specifically, for each combination of parameter values (table 2) correlations between simulated steady-state responses and empirical spatial activation profiles were calculated for inducer-up and inducer-down conditions. We selected those parameter values that produced largest average correlation under the additional constraint that spiking activation (unit output) reproduced the simultaneous brightness illusion. That is, larger activation in probe regions for the inducer-up condition compared to the inducer-down condition. Optimal model parameters were retained for the model comparison step.

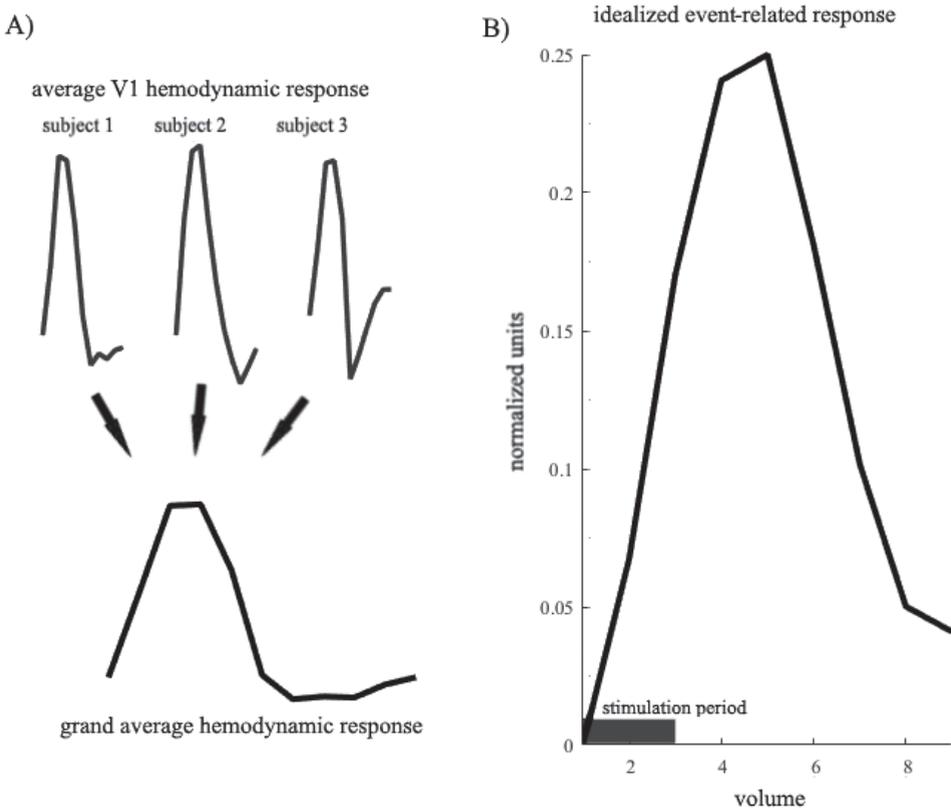


Figure 3.3: Idealized event-related response. *Panel A)* depicts the procedure by which we obtained the V1 grand average hemodynamic response. First, we calculated the average hemodynamic response of each subject over all V1 vertices. Subsequently, we averaged the resulting responses across subjects. *Panel B)* shows the idealized event-related response resulting from convolving the grand average hemodynamic response with the stimulation profile of each event. These stimulation profiles were, irrespective of condition, three volumes of stimulation followed by six volumes of baseline. The idealized event-related response was normalized to integrate to unity and reflects the expected shape and delay of observed event-related averages.

Model Comparison. We compared the two models based on their performance in accounting for the spatial activation profile observed in a test dataset (experimental run 2) using

model parameter values obtained from the previous step. Quantitative comparison of model performance was achieved using Akaike weights (Akaike, 1978, 1979) which indicate the conditional probability of each model. Calculations of these weights are based on the Akaike information criterion (AIC; Akaike, 1973) given by

$$AIC_i = -2 \log L_i + 2K_i \quad (3.1)$$

where $\log L_i$ reflects the log-likelihood of model i and K_i the number of free parameters in the model. Here, K was equal to five for both models. Absolute AIC values per model were transformed to relative AIC values, $\Delta_i(AIC)$ by subtracting the smaller of the two absolute values (Akaike, 1978). From these relative AIC values it is possible to obtain an estimate of the relative likelihood RL of each model i as

$$RL_i = e^{-\frac{1}{2}\Delta_i(AIC)} \quad (3.2)$$

Finally, Akaike weights of each model are obtained simply by dividing the relative likelihood of that model by the sum total of the relative likelihoods of both models (Wagenmakers & Farrell, 2004). We additionally compared models using the Bayesian Information Criterion (BIC; Schwarz et al., 1978). However, since these two criteria differ only with regard to the degree that they punish the inclusion of larger numbers of free parameters and since our models have the same number of parameters, results were identical for the two criteria and we only report comparisons based on the Akaike information criterion.

RESULTS

Empirical Results

Figure 3.4a shows each subject's psychophysical ratings of the strength and direction of the simultaneous brightness contrast effect expressed in normalized luminance (*black* = 0; *white* = 1) perceived in the probing regions for inducer-up and inducer-down conditions. Ratings for the inducer-up ($M = .37$, $SD = .04$) condition was lower than inducer-down ($M = .62$, $SD = .05$; $t[2] = 4.94$, $p = .0387$), indicating that subjects perceived the probing region to be darker when placed on a bright surround and brighter when placed on a dark surround.

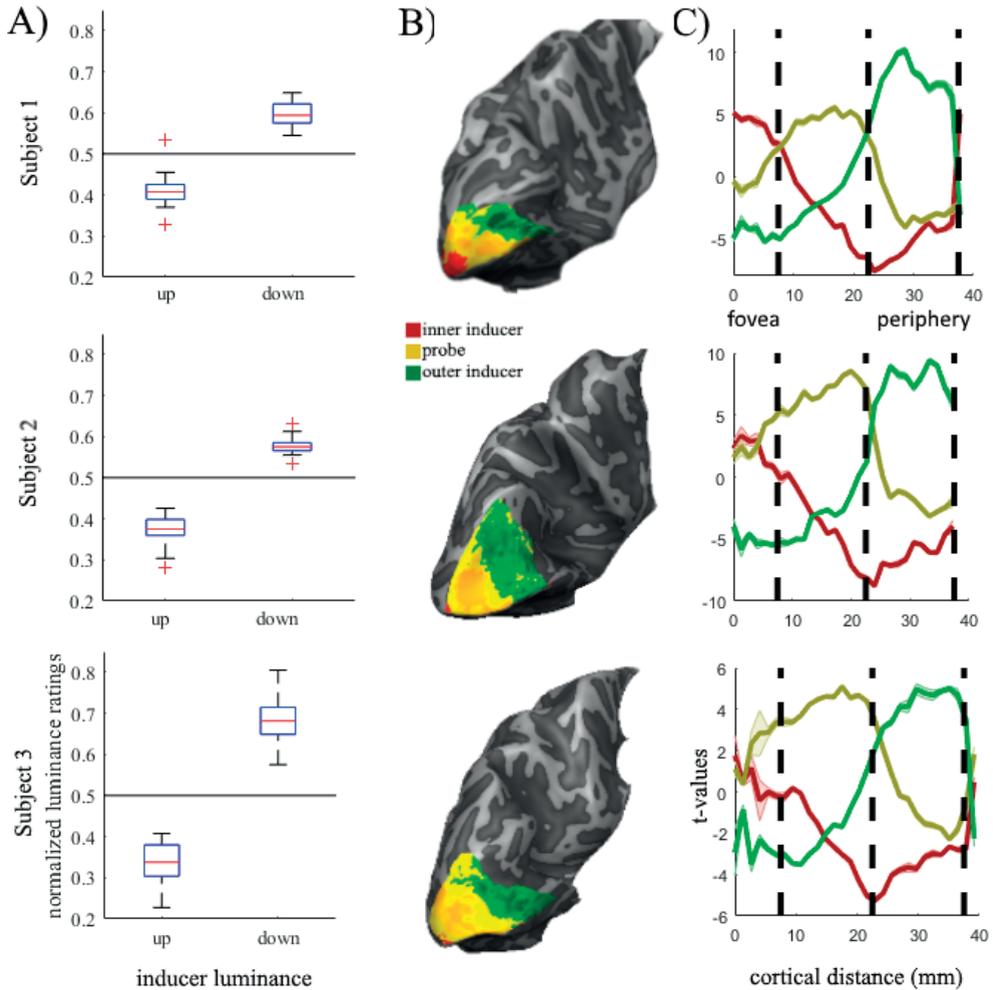


Figure 3.4: Psychophysical rating and localizer analysis. *Panel A)* shows the psychophysical brightness ratings (normalized with respect to the full range of luminance values) of our three subjects for the inducer-up and inducer-down condition of the simultaneous brightness experiment in the form of boxplots. The horizontal line indicates actual physical luminance values. On average subjects rate the probing region to be brighter than the physically present luminance in the inducer-down and darker in the inducer-up condition. *Panel B)* shows the results of the localizer contrast analysis overlaid on inflated cortex meshes of each subject's left hemisphere (results were similar for the right hemisphere). Red indicates vertices which were assigned to the inner inducer region while yellow and green indicate vertices which were assigned to the probe and outer inducer regions, respectively. *Panel C)* shows results of the localizer contrast analysis as a function of cortical distance for each subject. T-values of vertices falling within the same band were averaged (shaded regions represent SEM). Color code is identical to panel B). Dashed lines indicate the outer borders of the three rings.

Simultaneous brightness contrast fMRI data were subjected to group analyses with predictors modeling ROI-specific luminance, experimental condition, and their interaction. Analyses were performed on the inducer and probe ROIs, previously localized in each subject (figure

3.4b). In all regions, the absolute luminance covariate was significant and had the largest effect on the fMRI signal (table 3.1). However, in line with expectations and previous findings (van de Ven et al., 2011), there was also a significant interaction effect between condition and luminance which was most pronounced in the inner inducer and probe regions. Negative standardized beta coefficients for the interaction reflect a reduction in luminance signal in the experimental compared to the control condition. This indicates the presence of an additional signal in anti-phase to luminance changes in the experimental condition. Indeed, separate regression analyses per condition with relative and absolute luminance as predictors revealed positive beta coefficients for relative luminance in the control condition for the inner inducer ($\beta = .075, p \ll .01$) and probe ($\beta = .029, p = .0196$) regions, whereas the beta coefficients were negative in the same regions in the experimental condition (inner inducer: $\beta = -.045, p = .0174$; probe: $\beta = -.075, p \ll .01$). Beta coefficients were negative in both conditions for the outer inducer region ($\beta = -.030, p = .0131$; $\beta = -.094, p \ll .01$).

Table 3.1: Standardized beta coefficients for ROI analysis

	Inner	Probe	Outer
Intercept	.00	.00	.00
Luminance	.07*	.03*	-.03*
Condition	.00	.00	.00
Interaction	-.11*	-.10*	-.06*
Absolute luminance (covariate)	.16*	.11*	.18*

Results of the simultaneous brightness contrast group analysis conducted for the first experimental run for the three stimulus ROIs. Asterisks indicate significant results. Luminance, the interaction between luminance and condition, and the absolute luminance covariate were significant for each ROI. The significant negative interaction between luminance and condition indicates the presence of a signal in anti-phase to luminance changes was present in the experimental condition. The significant positive effect of absolute luminance indicates that the fMRI signal is partially driven by luminance changes per se irrespective of their direction (gray to white or gray to black).

In addition to the Stimulus-ROIs analyses, we analyzed fMRI responses as a function of cortical distance. Cortical-distance ROIs were computed by combining vertex-specific activity from each subject's two hemispheres and assigning resulting activity to one of 30 cortical distance bands (see Methods). For example, figure 3.4c shows subject-specific fMRI responses to the localizer stimuli as a function of cortical distance from the center of the stimulus (which corresponds to the locus of fixation). Likewise, figure 3.5 shows results for the interaction effect (figure 3.5a) and covariate (figure 3.5b) as a function of cortical distance. Note that the strongest negative interaction can be observed on the borders of inducer and probe regions (indicated by dashed lines). Furthermore, standardized beta coefficients observed in the probe region were generally more negative than those observed in either inducer region. Similar to the Stimulus-ROI analyses, absolute luminance change (included as covariate) has the largest effects in all areas. Interestingly, the covariate displays a trough within the probe region indicating that the effect observed in that region might be the result of spillover from the two inducer regions.

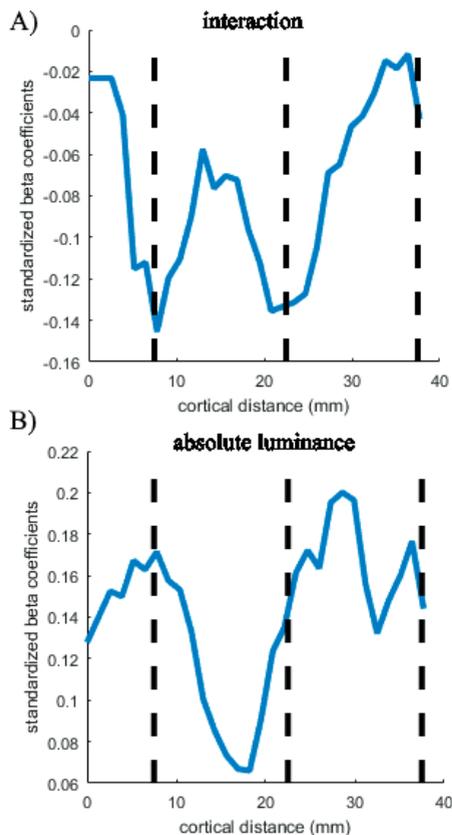


Figure 3.5: Simultaneous Brightness Contrast. Hemodynamic effects in Simultaneous Brightness Contrast task as a function of cortical distance (dashed lines indicate the outer borders of the three rings). *Panel A*) shows standardized beta coefficients resulting from the simultaneous brightness contrast group analysis for the interaction between condition and luminance. In the two inducer regions, standardized beta coefficients approach zero, whereas they are more negative in the probe region. Largest negative beta coefficients occur close to the borders between rings (dashed lines). *Panel B*) shows standardized beta coefficients for the absolute luminance change included as a covariate. Comparatively large beta values occur in the two inducer regions but drop considerably in the probe region.

Model Results

We integrated functional data from the first experimental run of the simultaneous brightness induction experiment in the common brain space (see figure 3.6) and used it for parameter optimization of the ODOG and EDGE model. Table 3.2 shows the parameter space that was explored for the two models as well as the parameter values resulting in the best average correlation between empirical and simulated spatial activation profiles over the two conditions (underlined values). Predicted spatial activation profiles of the two models for the inducer-up and inducer-down conditions given these parameter values are shown in figure 3.7. In line with studies showing that the fMRI signal largely (~80%) reflects synaptic rather than spiking activation (Logothetis et al., 2001; Maier et al., 2008), the weighting parameter in both models favors simulated synaptic over spiking activity. Specifically, the best fitting spatial activation profile produced by the ODOG model consisted to 33% of spiking and to 67% of synaptic activity. Similarly, the best fitting spatial activation profile produced by the EDGE model consisted to 22% of spiking and to 78% of synaptic activity. In evaluating the final correspondence between the best fitting model and empirical spatial activation profiles, we

opted for reporting resampling confidence intervals (based on 10000 samples of 10% of data points) because the large number of data points ($N = 797$ per condition) leads to overpowered parametric test statistics. The correlation between the empirical spatial activation profiles for inducer-up and inducer-down conditions and those predicted by the ODOG model were $r = .111$ (95% CI [-.077 .288]) and $r = .422$ (95% CI [.297 .538]), respectively. The correlation between empirical and predicted spatial activation profiles for the EDGE model were $r = .228$ (95% CI [-.021 .423]) and $r = .086$ (95% CI [-.068 .236]) for the inducer-up and inducer-down conditions, respectively. These results indicate that the EDGE model captures the inducer-up condition moderately well whereas the ODOG model captures the inducer-down condition very well.

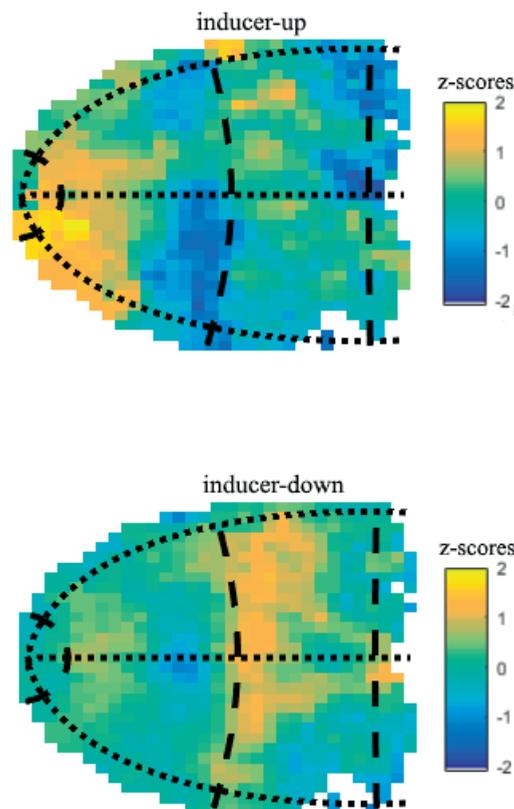


Figure 3.6: Empirical spatial activation profiles. The empirical CBS spatial activation profiles (peaks of event related averages per CBS location). The vertical and horizontal dotted lines represent the vertical and horizontal meridians, respectively. Dashed lines indicate, from left to right, the outer borders of the inner inducer, probe, and outer inducer stimulus regions. The upper row depicts the spatial profile (observed for the inducer-up condition while the lower row depicts the spatial profile observed for the inducer-down conditions).

Table 3.2: Model parameters

ODOG					EDGE				
λ	g	α	σ_o	W_o	σ_ϵ	b_1	b_2	b_3	W_ϵ
.20	.33	<u>.05</u>	<u>1.00</u>	.00	1.00	-1.75	-1.75	-1.75	.00
.23	.67	.10	1.33	.11	1.33	-1.5	-1.5	-1.5	.11
.27	1.00	.15	1.67	.22	1.67	-1.25	-1.25	-1.25	<u>.22</u>
.30	<u>1.33</u>	.20	2.00	<u>.33</u>	2.00	-1.15	-1.15	-1.15	.33
.33	1.67	.25	2.33	.44	<u>2.33</u>	-1	-1	-1	.44
.37	2.00	.30	2.67	.56	2.67	1	<u>1</u>	1	.56
.40	2.33	.35	3.00	.67	3.00	1.15	1.15	1.15	.67
.43	2.67	.40	3.33	.78	3.33	1.25	1.25	1.25	.78
<u>.47</u>	3.00	.45	3.67	.89	3.67	<u>1.5</u>	1.5	1.5	.89
.50	3.33	.50	4.00	1.00	4.00	1.75	1.75	<u>1.75</u>	1.00

Parameter space explored for the two Models. Parameters producing the best fit between model and empirical data are underlined and highlighted in bold. *ODOG model*. The parameter λ reflects the lowest of seven spatial frequencies (measured in cycles per mm) spaced at octave levels. We chose for the presented range of parameters since spatial frequency preference in human V1 peaks around .2 c/mm (H.-H. Yu et al., 2010). Parameters g and α controlled the weighting of spatial frequencies. The explored values remained close to those used in previous implementations of the ODOG model (Blakeslee et al., 2015; Blakeslee & McCourt, 1999, 2004). The slope α leads to the attenuation of low frequencies (Campbell & Robson, 1968; Georgeson & Sullivan, 1975). The parameter σ_o controlled the size of the Gaussian region for local normalization. We chose for the range given in the table because these values are close to the extent of lateral projections reported for macaque V1 (Angelucci et al., 2002). Finally, the parameter W_o controls the relative contribution of spiking and synaptic activity to the final activation profile. Since the fMRI signal is largely driven by synaptic activity (Logothetis et al., 2001; Maier et al., 2008), we desired to incorporate this in our model without specifying the exact contribution beforehand. *EDGE model*. The parameter σ_ϵ controlled the width of the Gaussian reflecting the extent of lateral impact of V1 interblob on V1 blob units. As for the other model, the explored values are close to the extent of lateral projections reported for macaque V1 (Angelucci et al., 2002). Three parameters, b_1 , b_2 , and b_3 , controlled the threshold in the logistic activation function. The explored values remained close to those used in previous implementations of the EDGE model (Peters et al., 2010). Finally, the parameter W_o controls the relative contribution of spiking and synaptic activity to the final activation profile.

Before evaluating the models against the test data set (experimental run 2), we assessed the adequacy of this approach by assuring that spatial activation profiles in the two experimental runs were indeed sufficiently similar. For the inducer-up condition the correlation was $r = .796$ (95% CI [.713 .855]) whereas it was $r = .875$ (95% CI [.830 .910]) for the inducer-down condition. Given the good correspondence between spatial activation profiles observed in the two runs we felt confident to proceed with model comparison using the Akaike information criterion. We again performed a resampling procedure and report median ratios of Akaike weights observed in 10000 samples of 10% of data points. For the inducer-up condition, the EDGE model ($\hat{W} = .974$) outperformed the ODOG model ($\hat{W} = .026$) as it was 37.57 times more likely to be the best model for this dataset. The correlation between the inducer-up spatial activation profile predicted by the winning (EDGE) model and that observed in the second empirical run was $r = .197$ (95% CI [-.059 .398]). For the inducer-down condition, the ODOG ($\hat{W} > .999$) model strongly outperformed the EDGE model ($\hat{W} < .001$) as it was more than 10000 times

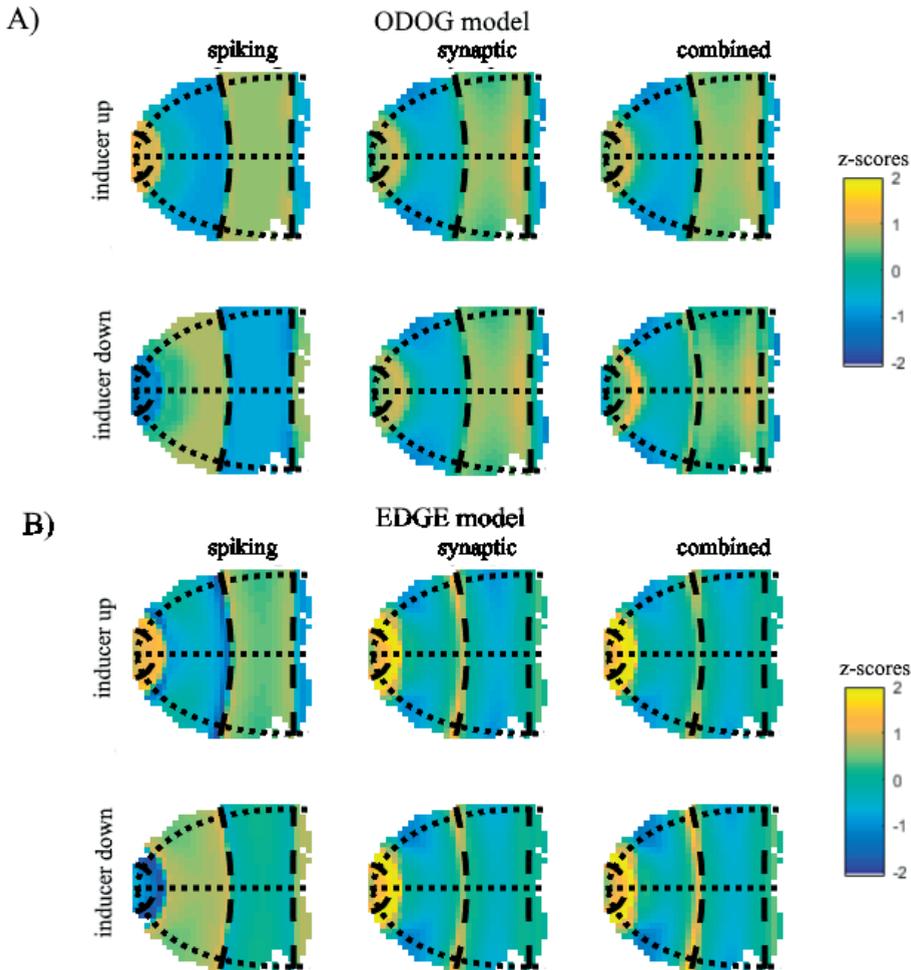


Figure 3.7: Model spatial activation profiles. This Figure shows z-normalized CBS spatial activation profiles of the ODOG and EDGE model in *panels A) and B)*, respectively. The left column shows predicted spiking (unit output) activity, the middle column shows predicted synaptic (unit input) activity, and the right column shows the weighted combination of the two. The top row of each panel shows predictions for the inducer-up condition while the lower row shows predictions for the inducer-down condition. Dotted lines represent the horizontal and vertical meridians defining the coordinate system of the common brain space. Dashed lines represent the outer borders of the inner inducer, probe, and outer inducer regions. In accordance with the constraints imposed during the fitting procedure, spiking activation profiles reflect the simultaneous brightness contrast percept (see methods) and were highly similar for the two models. Correlations between the two were $r = .886$ and $r = .814$, for the inducer-up and inducer-down conditions, respectively. Synaptic activation profiles were less similar with $r = .198$ and $r = .429$ for the two conditions, respectively. Since our parameter estimation procedure suggested that the best predictions of fMRI spatial activation profiles result from a bias towards synaptic activity in their combination, these lower similarities are advantageous for model discriminability. Correlations between best fitting model spatial activation profiles (combining spiking and synaptic activity) were $r = .427$ and $r = .421$ for the two conditions, respectively. Best fitting spatial activation profiles of the two models differ largely with respect to the activation they predict at borders between rings and within the outer inducer region.

more likely to be the best model for this dataset. The correlation between the inducer-down spatial activation profile predicted by the winning (ODOG) model and that observed in the second empirical run was $r = .411$ (95% CI [.284 .521]).

DISCUSSION

We developed a procedure to compare the prediction accuracy of computational models with respect to spatial activation patterns in early visual cortex as observed with ultra-high-field fMRI. A key element of this procedure is a functionally defined common brain space wherein empirical and simulated data can be represented in the same coordinate system. Data represented within this space could be analyzed using established statistical tools in fMRI research. The real strength of the CBS is, however, that it enables the evaluation of competing computational models in terms of their aptitude to account for the spatial activation profile of empirical data. For this evaluation benchmarks such as the Akaike or the Bayesian information criterion (Akaike, 1973; Schwarz et al., 1978) can be used. We exemplified this procedure in the context of simultaneous brightness contrast. First, we reproduced results showing that the fMRI signal in early visual cortex reflects the simultaneous brightness contrast effect (van de Ven et al., 2011). We found that a signal in anti-phase to physical luminance change was present in V1 if the probe region was gray but not when it was black (control condition). This anti-phase signal was strongest at region borders and stronger inside the probe region relative to the inducer regions.

These results confirm that our data indeed contained relevant information with regard to the simultaneous brightness contrast effect, a necessary prerequisite for model comparison. Consequently, we compared predictions of a neural implementation of an oriented difference-of-Gaussian model to a neural network model involving an edge integration mechanism, based on their similarity to fMRI activity observed in our 7T simultaneous brightness contrast experiment. Comparisons were performed for static spatial profiles of fMRI activity related to an inducer-up and an inducer-down condition. For each model both simulated spiking (unit output) and synaptic (unit input) activity contributed to the predicted fMRI activation profiles. The weighting between the two, as obtained in a parameter fitting procedure, was 33% spiking plus 67% synaptic for the ODOG model and 22% spiking plus 78% synaptic for the EDGE model. These weightings were close to the relative contributions reported previously where spiking contributed about 20% and (subthreshold) synaptic activity about 80% to the fMRI signal (Logothetis et al., 2001; Maier et al., 2008). Given these and the remaining parameter values, the EDGE model outperformed the ODOG model for the inducer-up condition of our dataset. According to the Akaike comparison it was around 37 times more likely to be the better of the two models. For the inducer-down condition, the ODOG model strongly

outperformed the EDGE model since it was around 10000 times more likely to be the better of the two models according to the Akaike comparison.

These relative outcomes are reasonable given that the EDGE model produced a (moderate) fit with the inducer-up condition only whereas the ODOG model produced a (good) fit with the inducer-down condition only. The observation that each model can only account for a single of the two conditions stems from the fact that both models in effect produce a single synaptic activation profile regardless of condition in conjunction with the strong weighting of synaptic activity for the overall prediction of spatial activation profiles. The former is due to synaptic activity reflecting absolute input and due to highly symmetric mechanisms in both models producing inputs (as well as outputs) that differ in sign but not magnitude. Our results thus show that the underlying mechanism accounting for simultaneous brightness contrast must also account for differences in input (output) magnitude depending on the condition and hence contain asymmetric aspects. That the EDGE model specifically reproduces the fMRI spatial activation profile observed in the inducer-up but not in the inducer-down condition might be due to an imbalance in lateral excitation and inhibition spreading from the edges with a dominant inhibitory effect. This would be in line with findings that suppressive, but not facilitating, lateral connectivity can exhibit spatially asymmetric profiles (Walker et al., 1999). The EDGE model assumes that a balanced interaction between symmetric facilitating lateral connectivity and asymmetric suppressive lateral connectivity produces excitatory and inhibitory modulation of equal magnitude spreading in opposite directions. However, the assumed balance appears not to be supported by our results.

It is too early to draw general conclusions with regard to the mechanism accounting for surface perception based on our results since we only compare very specific model implementations of each. Different implementations might perform entirely differently. To evaluate the overarching mechanisms one would need to target essential aspects of each that would be shared across implementations. Nevertheless, our results point to the possibility that early visual cortex might utilize a spatial filtering as well as an edge integration mechanism when processing surface luminance. Specifically, an asymmetric edge integration process wherein edges predominantly suppress activation in the darker of two adjacent regions might complement a symmetric spatial filtering process. To elucidate this question, the common brain space might be used to develop a biologically plausible model of such a hybrid mechanism as it would allow for the comparison of an array of candidate implementations. Furthermore, each design choice could immediately be evaluated in terms of its implications with respect to the spatial activation profile.

The adequacy of the CBS approach to theory-driven fMRI research is strongly dependent on the quality of the mapping of spatial and temporal aspects of (simulated) neuronal signals to fMRI signals. In turn, the quality of spatial mapping relies to a large degree on how well the cortical coordinates of vertices (or voxels) in early visual cortex can be measured and hence on choices with respect to population receptive field estimation. We have used the standard

isotropic 2D Gaussian pRF model in conjunction with recommended stimulation procedures (Dumoulin & Wandell, 2008; Senden, Reithler, et al., 2014) resulting in good representations. However, estimation of vertices' cortical coordinates could be improved beyond what was reported here by taking cortical magnification into account for stimulus design and search space definition. Particularly mapping of foveal voxels would benefit from these adaptations. Further improvements to the common brain space approach will also depend on increased understanding of the relation between neural activation and the spatial and temporal aspects of the hemodynamic response (Havlicek et al., 2015; Logothetis et al., 2001; Shmuel, Yacoub, Chaimow, Logothetis, & Ugurbil, 2007).

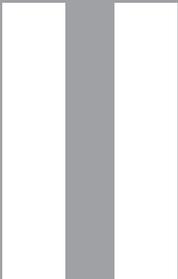
In conclusion, our result show that a functionally defined common brain space can indeed provide a suitable reference frame for the evaluation of models' predictions of spatial activation profiles. With regard to surface perception our model comparison results suggest a hybrid mechanism. Furthermore, the fact that the common brain space can be used to evaluate candidate implementations against each other, suggests that the common brain space approach is useful not just for theory-testing but for theory-building as well.

ACKNOWLEDGEMENTS

This project was supported by the European Research Council under the European Union's Seventh Framework Programme (ERC-2010-AdG, grant agreement number: 269853). We thank Dr. Dimo Ivanov for his help while setting up the used MR imaging protocols.

Theory Building

Part



Chapter

4

Rich Club Organization Supports a Diverse Set of Functional Network Configurations

This chapter is an adapted version of the following article:

Senden, M., Deco, G., de Reus, M. A., Goebel, R., & van den Heuvel, M. P. (2014). Rich club organization supports a diverse set of functional network configurations. *NeuroImage*, 96, 174–182

ABSTRACT

Brain function relies on the flexible integration of a diverse set of segregated cortical modules, with the structural connectivity of the brain being a fundamentally important factor in shaping the brain's functional dynamics. Following up on macroscopic studies showing the existence of centrally connected nodes in the mammalian brain, combined with the notion that these putative brain hubs may form a dense interconnected 'rich club' collective, we hypothesized that brain connectivity might involve a rich club type of architecture to promote a repertoire of different and flexibly accessible brain functions. With the rich club suggested to play an important role in global brain communication, examining the effects of a rich club organization on the functional repertoire of physical systems in general, and the brain in particular, is of keen interest. Here we elucidate these effects using a spin glass model of neural networks for simulating stable configurations of cortical activity. Using simulations, we show that the presence of a rich club increases the set of attractors and hence the diversity of the functional repertoire over and above the effects produced by scale free type topology alone. Within the networks' overall functional repertoire rich nodes are shown to be important for enabling a high level of dynamic integrations of low-degree nodes to form functional networks. This suggests that the rich club serves as an important backbone for numerous co-activation patterns among peripheral nodes of the network. In addition, applying the spin glass model to empirical anatomical data of the human brain, we show that the positive effects on the functional repertoire attributed to the rich club phenomenon can be observed for the brain as well. We conclude that a rich club organization in network architectures may be crucial for the facilitation and integration of a diverse number of segregated functions.

INTRODUCTION

The human brain is composed of a large set of anatomically distinct regions and local clusters indicative of segregated neural information processing. The execution of higher order cognitive functions such as memory, perception, and attention demands the integration of this information via distributed computation. Computationally driven theories of cognition hypothesize that the brain may achieve integration of subsystems by flexibly arranging cortical areas into temporal functional networks in accordance with goal-related requirements (Baars, 2005; Deco, Jirsa, & McIntosh, 2010; Ghosh, Rho, McIntosh, Kötter, & Jirsa, 2008). The exact nature as well as the size of the set of possible functional network configurations, referred to as the brain's functional repertoire, has been suggested to relate directly to the structural architecture of the brain (Deco et al., 2010; Deco, Senden, & Jirsa, 2012; Senden, Goebel, & Deco, 2012). Network architectures that involve a scale free topology; meaning that the degree distribution follows a power law function indicating the existence of a small number of high-degree nodes, have been shown to be able to display a particularly diverse number of functional configurations (Deco et al., 2012).

In addition to a heavy tailed degree distribution the human brain has been shown to contain hubs which are not only individually 'rich' in connectivity but additionally show a dense level of interconnectivity (Colizza et al., 2006; van den Heuvel & Sporns, 2011). This collective of highly interconnected hubs has been termed the 'rich club' analogously to the organization of social systems in which individuals rich in connections tend to form strongly interconnected clubs, taking a central position in the overall system (McAuley, da Fontoura Costa, & Caetano, 2007; Zhou & Mondragon, 2004). Similarly, neural rich clubs have been hypothesized to act as a central high-capacity backbone for global communication (van den Heuvel, Kahn, Goñi, & Sporns, 2012) and integration (van den Heuvel & Sporns, 2013a) in the brain.

In this computational study we hypothesized that the presence of a rich club in an otherwise scale free type architecture further expands the functional repertoire of a system. To test this hypothesis a steady-state attractor model was applied to artificial network architectures as well as human experimental anatomical data to examine the influence of a network's architecture on its functional diversity. In addition, we studied the shape of the functional repertoire with a special focus on the comparison between scale free architectures that show rich club organization versus scale free architectures without rich club organization. Overall, our findings suggest that the presence of a central rich club on top of a scale free architecture may lead to an additional gain in the diversity of a network's functional repertoire, suggesting that a structural rich club in neural systems may further expand the brain's functional diversity. Interestingly, our findings further indicate that the presence of a rich club enhances the capability of the system to functionally integrate areas of low structural degree.

METHODS

Artificial Network Architectures

Six types of network architectures were examined (figure 4.1): Regular (REG), Random (RAND), Small world (SW), Barabási-Albert (Barabási & Albert, 1999) scale free including a rich club formation (SF-RC), Scale free without rich club formation (SF), and Scale free whose rich nodes are secluded from one another (SF-negRC). Each network contained exactly 24 nodes (results of networks of $N=30$ nodes are presented in the supplemental materials) with each node, on average, making four bidirectional, unweighted connections to other nodes, resulting in a total of 96 connections (also referred to as edges) per network. Small networks were chosen due to the high computational demand of the spin glass model (see supplemental information). In what follows, the formation of these networks are described, starting with the REG, RAND and SW class, followed by the three categories of SF networks.

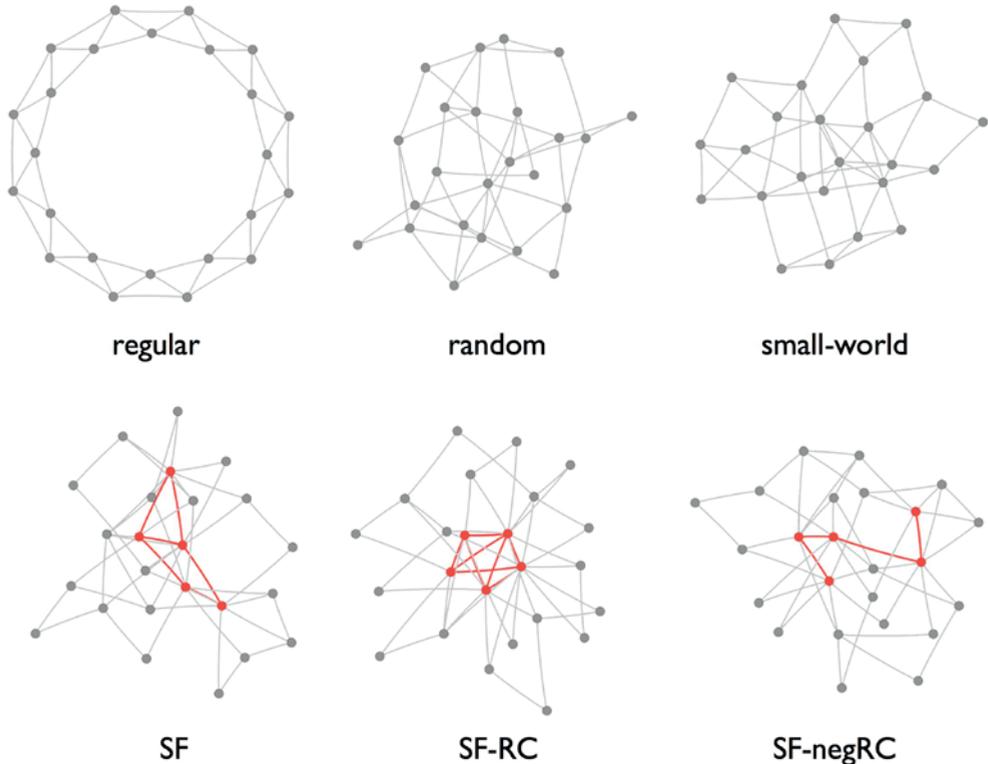


Figure 4.1: Network architectures. The figure gives an illustrative representation of the artificial network architectures studied here. The upper row shows a regular (named the REG class, left), small world (SW, middle), and random (RAND, right) network. The lower row shows three scale free networks with different rich club organizations. Left is a typical scale free network without rich club organization (SF class). In the middle is the scale free network showing rich club organization (named the SF-RC class). Right is a scale free network showing negative rich organization (SF-negRC class). High degree hub nodes and hub-to-hub connections are shown in red.

Regular network (REG). A single regular network was generated by ordering 24 nodes on a circular lattice and subsequently connecting each node to its two nearest neighbors on both sides.

Random networks (RAND). A set of 100 random networks were generated using the algorithm described by Watts and Strogatz (1998) which rewires each connection in a regular network with a prefixed probability p . For random networks the probability of rewiring was set equal to one.

Small world networks (SW). A set of 100 small world networks were generated using the rewiring algorithm described by Watts and Strogatz (1998) with a probability of rewiring set to 0.25, resulting in a network with a small world topology in which the majority of edges are between neighboring nodes, with a few connections forming short-cut connections between remote parts of the network.

Scale free networks. Scale free networks were generated by applying the Barabási-Albert algorithm (Barabási & Albert, 1999) on random seed networks. This algorithm employs the principle of preferential attachment in which the probability that a newly added node will form a connection with an existing node is proportional to the degree of the existing node. As a result, the subset of seed nodes will end up as the most densely connected nodes (i.e. hubs) in the generated network. Three types of scale free networks were formed:

1. Scale free networks without a rich club (SF). A set of 100 scale free networks containing no rich club (SF) were generated by performing the Barabási-Albert algorithm (Barabási & Albert, 1999) on a seed of intermediate density. The seed set had nine nodes each making four connections leading to the formation of a set of scale free networks in which the starting nodes show a high level of connectivity (i.e. form hubs) but no central rich club.
2. Scale free networks with a rich club (SF-RC). A set of 100 scale free networks containing a dense rich club (SF-RC) were generated by performing the Barabási-Albert algorithm (Barabási & Albert, 1999) on a dense seed set of six nodes each making four connections, leading to the formation of a set of scale free networks in which the starting nodes show a high level of connectivity (i.e. form hubs) as well as a dense level of interconnectivity forming a central rich club.
3. Scale free networks with a negative rich club. A set of 100 scale free networks and with a negative rich club (SF-negRC) was generated by performing the Barabási-Albert algorithm (Barabási & Albert, 1999) using a sparse seed of twelve nodes with each making four connections, resulting in a set of networks with a topology that showed a scale free degree distribution, but in which the hubs had a below chance level of inter-connectivity (i.e. formed a below chance level of connectivity).

Spin Glass Model

To examine the link between network architecture and functional entropy we adopted the analytically solvable Ising spin glass model from Deco *et al.* (2012). The model, which is isomorphic to the discrete Hopfield net (Hopfield, 1982), studies the characteristics of the attractor landscapes emerging in a spin glass neural model. The spin glass model includes a network of spins which can be in one of two possible states (0,1). The spins are symmetrically coupled according to a set of connections (here, the undirected binary connections are edges given by an underlying structural network and the spins represent cortical areas), allowing for interactions among them. The state of a spin i is given by S_i , C denotes the adjacency matrix associated with the network architecture under consideration. The probability of finding the network in a specific global configuration S (indicated by a superindex α) is derived from the Boltzmann-Gibbs distribution (Gibbs, 1905)⁵:

$$P^\alpha = \frac{e^{-\varepsilon H^\alpha}}{Z} \quad (4.1)$$

where ε is the reciprocal of the thermodynamic temperature of the system (here, $\varepsilon = 1$) and Z is the partition function

$$Z = \sum_{\alpha} e^{-\varepsilon H^\alpha} \quad (4.2)$$

The energy function H^α is given by

$$H^\alpha = \frac{1}{2} \theta \sum_i S_i^\alpha - \frac{1}{2} G \sum_{i,j} C_{ij} S_i^\alpha S_j^\alpha \quad (4.3)$$

with θ being the transition threshold at which a spin changes its state and G a scaling parameter for the adjacency matrix C referred to as global coupling strength. Here, θ was set to 12. The entropy of the system is given by:

$$E = -\sum_{\alpha} P^\alpha \log P^\alpha = \frac{\sum_{\alpha} \varepsilon H^\alpha e^{-\varepsilon H^\alpha}}{Z} + \log Z \quad (4.4)$$

and reflects the number of accessible states of the system and serves to describe the attractor landscape of the system. Note that the maximal entropy that a system with N binary nodes can exhibit is given by:

⁵ A more recent English description is given in (Landau & Lifšic, 2007)

$$\max(E_N) = \log(2^N) \quad (4.5)$$

Solving the spin glass model, which requires examination of 2^N states, is computationally demanding even for small networks (see figure A. in the appendix A5.1). In order to investigate spin glass models with up to 2^{30} states the energy function as well as the summands of the partition function were calculated for all states in parallel on graphical processing units (GPUs). To be able to investigate spin glass models with more than $N = 30$ nodes and 2^{30} states we adopted a Metropolis-Hastings Algorithm (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953), sampling a subset out of the total 2^N solutions (see appendix A4 for details).

We additionally derived the mutual information from the probability distribution P^{θ} across states, which allows for calculation of both joint and marginal probabilities. Specifically, the mutual information M between nodes i and j is given by:

$$M_{i,j} = \sum_{S_i=0}^1 \sum_{S_j=0}^1 P(S_i, S_j) \log \left(\frac{P(S_i, S_j)}{P(S_i)P(S_j)} \right) \quad (4.6)$$

where $P(S_i)$ and $P(S_j)$ are the marginal probabilities of nodes i and j , respectively, and $P(S_i, S_j)$ are the joint probabilities of the two nodes.

Degree-Related Connectivity

Degree-related connectivity (DRC) was used to quantify the strength with which nodes of a certain degree range are interlinked. More formal, it measures the total observed connectivity between nodes of degree k and nodes of degree l divided by the number of possible structural connections between those nodes, with k and l being structural degrees obtained from the network architecture.

$$DRC_{k,l} = \frac{\sum_{i \in N_k} \sum_{j \in N_l} A(i, j)}{\sum_{i \in N_k} i \sum_{j \in N_l} j} \quad (4.7)$$

where i and j are nodes from the sets N_k and N_l comprising all nodes of degree k and l within the underlying adjacency matrix, respectively and A is the connectivity under consideration. In the case that nodes of equal degree are considered (i.e. k is equal to l) both i and j are drawn from the same population N_k . Here, potential self-connections are omitted by drawing j from N_k and i from $N_k \setminus \{j\}$. For structural DRC A is identical to the adjacency matrix C . To compute *functional* DRC (i.e. DRC based on functional connectivity) A was identical to a functional connectivity matrix given by the mutual information matrix MI (or by the cross-correlation matrix for empirical data). Division by the number of possible structural connections between nodes

of certain degrees is necessary since certain degrees can be overrepresented in network architectures and hence bias the DRC. This is especially true for scale free architectures where low-degree nodes are overabundant.

Human Empirical Data

In addition to the created artificial network architectures (i.e. regular, random, small world, and three types of scale free networks of 24 nodes) we investigated the behavior of the attractor model on empirical human anatomical connectivity data, derived from diffusion weighted imaging. Group-averaged structural data and information on the brain's functional connectivity architecture were acquired on the basis of diffusion weighted MRI and resting-state fMRI recordings, respectively, based on data described in the recent paper of Collin *et al.* (2013). In short we refer to Collin *et al.* (2013) for a detailed description of the acquisition and analysis procedure), the selected subset described the reconstructed anatomical connections between 68 areas of the cortex of 23 healthy adult subjects derived from deterministic streamline tractography. Next, a binary group-average *structural connectivity* matrix was formed by including all connections that were found to be present in at least 70 percent of the participants (de Reus & van den Heuvel, 2013). In parallel, a group-averaged *functional connectivity* matrix was formed by averaging the individual functional connectivity matrices over the group of subjects, describing the level of correlation between the resting-state time-series of each pair of the included 68 cortical regions. Structural DRC was calculated for empirical structural connectivity matrices while functional DRC was calculated for empirical functional connectivity matrices. All empirical connectivity matrices comprised 68 nodes (i.e. both hemispheres). Computation of empirical DRC was performed in a similar fashion as computed for the artificial networks. However, computation of functional DRC was based on the group-averaged functional correlation matrix (68 nodes); taking into account information on the positive correlations (i.e. negative correlations were ignored). The procedure to calculate functional DRC remained otherwise identical to the one used for *MI* in the artificial networks.

RESULTS

Rich club organization

The three categories of scale free networks were quantitatively distinguished by their normalized rich club coefficients given by the fraction of the actual number of connections between nodes with degree larger than k to the maximal number of connections among these nodes (Colizza *et al.*, 2006; Zhou & Mondragon, 2004). Rich club coefficients were calculated for each of the scale free networks and normalized to a set of random networks with equal degree distributions with a normalized rich club coefficient exceeding 1 signifying the presence of rich club organization within a network (Collin *et al.*, 2013; van den Heuvel & Sporns, 2013b).

Figure 4.2 shows the rich club coefficients for the three classes of scale free networks. Validating our network generation models, the SF-RC class revealed an average normalized rich club function that exceeds 1 in the interval [4,7] (dark blue line), the SF class (blue line) showed an average rich club coefficient close to 1, and the SF-negRC class (light blue) showed an average rich club coefficient below 1.

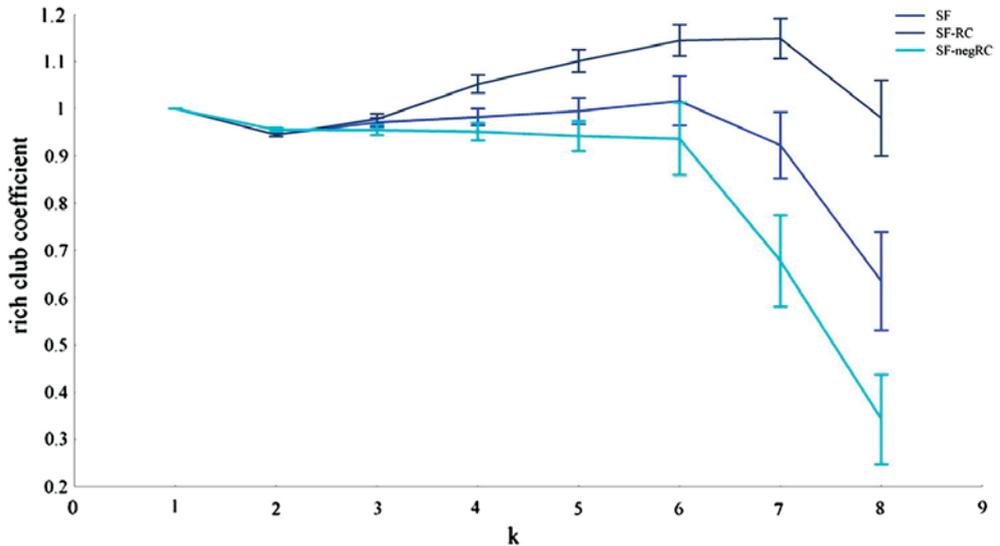


Figure 4.2: Normalized rich club coefficient. The figure shows the normalized rich club coefficient given as a function of degree for the three types of scale free networks. As expected from the manipulation procedures, SF-RC networks exhibit a mean rich club coefficient exceeding 1 in the interval [4,7] (dark blue line). Normal SF networks in which hubs do not form a rich club exhibit a mean rich club coefficient around 1 (blue line); SF-negRC networks whose high-degree nodes show low levels of interconnectivity, exhibit a mean rich club coefficient consistently smaller than 1 (light blue line).

Entropy

The entropy of each of the network models was examined by using a spin glass model in which spins are arranged on a lattice according to an underlying structural architecture and connection strengths are systematically increased by a global coupling factor. At critical coupling the system as a whole becomes multistable and exhibits a set of attractors representing the functional repertoire of the system, with the system's level of entropy reflecting the size of the attractor set (Deco et al., 2012). Figure 4.3 shows the entropy of the attractors for the set of networks as a function of the global coupling strength for all classes of artificial networks. Supporting previous findings (Deco et al., 2012), the entropy obtained across a range of global coupling strength was found to be higher for the scale free network architectures as compared to SW, RAND, and REG (for an analysis with networks comprising 30 nodes see figures A5.2 and A5.3 in the appendix). However, importantly, being the focus of our study, within the class

of scale free architectures, networks of the SF-RC class (i.e. containing a rich club) revealed a higher level of entropy (mean entropy $\mu E = 10.23$) than networks of the SF ($\mu E = 9.54$) and SF-negRC ($\mu E = 8.79$) class [$p < .001$ in an ANOVA test with entropy at the critical coupling point ($G = 3$) as dependent variable and rich club (rich club, no rich club, negative rich club) as fixed factor]. The largest theoretically possible entropy among network architectures comprising 24 nodes was computed as $\max(E_{24}) = 16.636$.

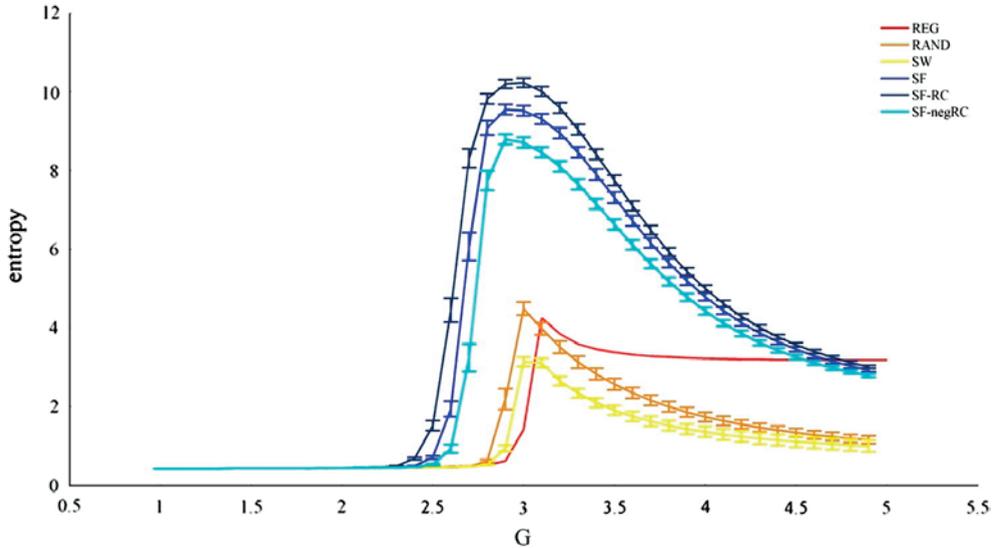


Figure 4.3: Entropy of the attractors of spin glass networks (24 nodes). The figure shows the entropy observed for each network class as a function of global coupling strength. All networks included the same network size and connectivity density (twenty-four nodes and 96 edges). The scale free networks show larger entropy levels over attractors than the corresponding small world (SW, yellow), regular (REG, red) and random (RAND, orange) networks. Additionally, within the category of scale free networks, the class of networks exhibiting rich club organization (SF-RC networks, dark blue) display the largest levels of entropy, while those exhibiting negative rich club organization (SF-negRC, light blue) show the lowest levels of entropy. Scale free networks without a rich club (SF, blue) show an intermediate level of entropy.

Functional Repertoire

The marginal probabilities of nodes being active, i.e. their corresponding spins being in an ‘up’ state, give an indication as to how many nodes are active at any moment which limits the number of possible configurations. The largest number of possible configurations can be achieved by systems with 50% of their nodes active for any given configuration. Interestingly, the scale free network architectures considered here approached the 50% mark. In detail, for SF network architectures the most common global configurations were those where 65.5 percent of the nodes were active. However, for the SF-RC network architectures the most common global configurations were those where 58.3 percent of the nodes were active, higher (and therefore more diverse) as compared to the SF networks. Finally, for SF-negRC

network architectures the most common global configurations were those where 66.7 percent of the nodes were active. For comparison, for regular, random, and small world network architectures the most common global configurations were those where, either all or none, 79.2 percent, and 70.8 percent, of the nodes were active, respectively.

While these numbers represent the overall number of nodes being active for the most configurations, the distribution across nodes is not uniform. Specifically, high-degree nodes in scale free network architectures revealed a particularly high probability of being active in such a way that their spins were in an 'up' state for all of the common global configurations. At the same time low-degree nodes in these network architectures were active for 50% or less of the common global configurations. This is illustrated in figure 4.4 where the marginal probability of being active is shown for each of the 24 nodes within a network architecture with nodes sorted ascending by their degree (random network architectures served as a baseline). Rich nodes (i.e. the set of high degree nodes) revealed to be active almost unequivocally, whereas low degree nodes were far freer to switch between 'up' and 'down' states. Although this effect was present for all scale free type architectures, it was found to be the most pronounced for SF-RC architectures.

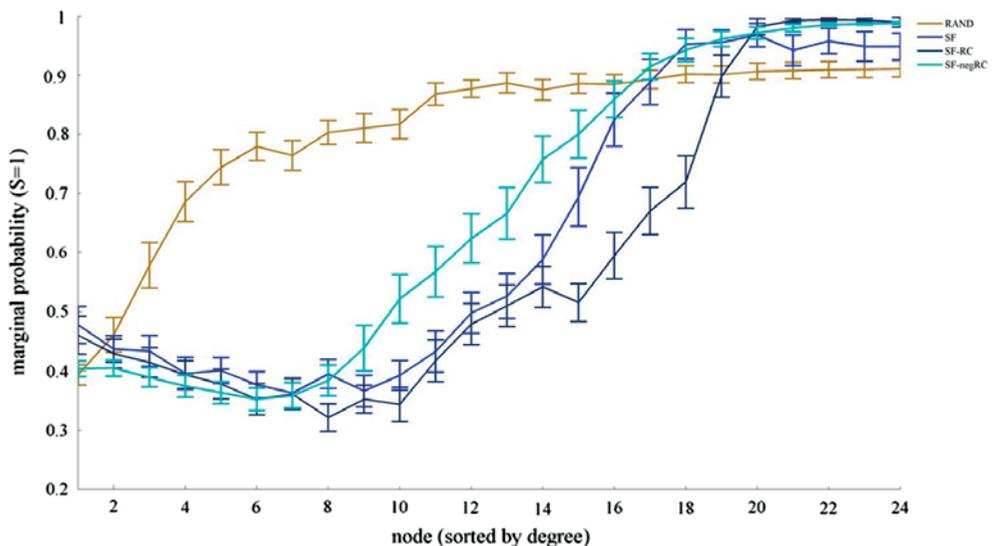


Figure 4.4: Marginal probability of 'up' states. The figure shows the marginal probability of being in an 'up' state for each node of a network architecture. The nodes are sorted by their degree. For all networks the marginal probability rises with degree. For random networks (orange) the rise resembles a logarithmic curve. In contrast, for scale free network architectures the marginal probability curve resembles an 'S' curve with regions of low and high probability separated by a sharp rise. For SF-RC networks (dark blue) the region of high probability is smaller as compared to the other scale free type architectures. Additionally, for SF-RC networks two regions exist where nodes exhibit a lower probability of being active as compared to SF (blue) and SF-negRC (light blue) networks. These findings indicate that for scale free network architectures high-degree nodes contribute to every functional network while more low-degree nodes differentially contribute to different functional networks. For SF-RC networks more nodes were free to contribute to several functional repertoires.

The high degree - low degree subdivision of the functional repertoire was further examined using degree related connectivity. Figure 4.5a and b show the average structural and functional DRCs over the 100 matrices of SF, SF-RC and SF-negRC architectures, with figure 4.5c showing the Euclidean distance between the centers of mass of functional and structural DRC distributions. The structural connectivity for SF network architectures largely featured connections among nodes of moderate degree with the mean center of mass (which in symmetric distributions such as these lies along the diagonal) over all structural DRCs at $\mu_k = 6.73$ 95% CI [6.62 6.83]. As expected from our manipulations, for SF-RC the mean center of mass over all structural DRCs shifted toward higher degrees $\mu_k = 7.19$ 95% CI [7.08 7.3]. For SF-negRC the mean center of mass over all structural DRCs was located at $\mu_k = 6.29$ 95% CI [6.19 6.39]. With regard to functional DRC, low degree functional connections were found to be prominent for SF-RC network architectures but less prominent in SF and SF-negRC network architectures as the mean centers of mass were located at $\mu_k = 5.25$ 95% CI [5.14 5.35], $\mu_k = 6.15$ 95% CI [6.04 6.25], and $\mu_k = 6.21$ 95% CI [6.12 6.3], respectively. These differences among center of mass for functional DRCs were found to be significant in an ANOVA test with $p < .001$. Subsequent pairwise comparison revealed that the center of mass for SF-RC was located at significantly lower degree nodes as compared to SF ($p < .001$) and SF-negRC ($p < .001$), whereas no significant difference was observed between the centers of mass of the latter two ($p = .34$). Taken together, these findings suggests that while SF and SF-negRC mainly recruit nodes of moderate degree for the formation of functional networks, a SF-RC architecture may enable the formation of a diverse set (and overall, a more diverse set as reflected by the higher level of entropy, figure 4.3) of functional configurations recruiting low-degree nodes. Since low-degree nodes are most prominent in scale free network architectures (see figure A5.4 in the appendix for details) their integration allows for a particularly diverse functional repertoire.

The third column of figure 4.5 shows Euclidean distances between centers of mass for functional and structural DRC, showing that functional interactions can be strong in the absence of structural connections. The distribution of Euclidean distances was obtained by a bootstrapping procedure sampling the structural and functional centers of mass from the 100 networks considered for each class of network architectures. Figure 4.5c emphasizes that SF-RC networks show strong connectivity among low-degree nodes even though structural connections were mostly centered on moderate to high-degree nodes. For SF and SF-negRC network architectures the formation of functional connections was found to be more limited to nodes which already share anatomical connections.

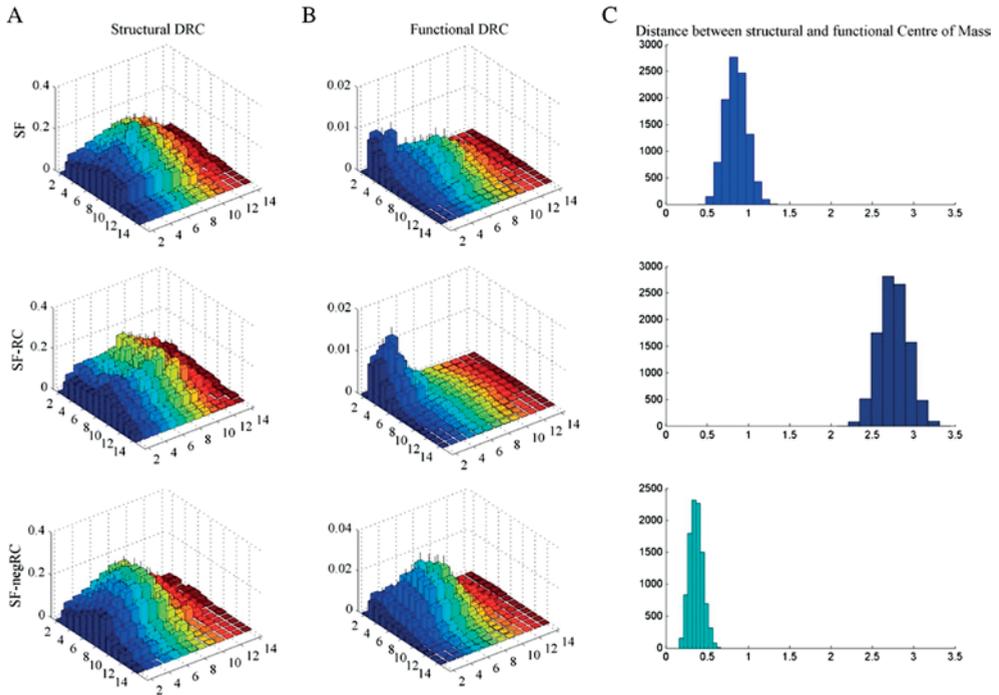


Figure 4.5: Degree-related connectivity in artificial network architectures. The figure shows degree related connectivity for structural and functional connectivity derived from simulations of the three classes of scale free network architectures. For ease of interpretation the level of degree was divided into three categories, being low ($k = [2,5]$), moderate ($k = [6,10]$) and high ($k = [11,14]$). *Panel A*) The first column shows average structural degree-related connectivity (DRC) for all types of network architectures. For SF networks (top row) connections between moderate and high-degree nodes are most prominent. As expected in the absence of a rich club, high-to-high connections are rare. For SF-RC networks (middle row) moderate-to-high connections are again most prominent, but in contrast to SF networks, networks with a SF-RC architecture feature a significant number of high-to-high connections. For SF-negRC networks (lower row) moderate-to-moderate connections are most prominent. Finally, in-degree connections among the most prominent nodes, namely those exhibiting low degree (76% of the nodes in the networks, see also figure A5.4 in the appendix), are largely absent for all three scale free network architectures. *Panel B*) The second column shows functional DRC for all scale free network architectures. For SF networks (upper row) moderate-to-moderate connections are most common, but they also feature some low-to-low and low-to-moderate connections. For SF-RC networks (middle row) low-to-low interactions are most common. For SF-negRC networks (lower row) the main concentration of connections is centralized among nodes of moderate degree. *Panel C*) The third column shows Euclidean distances between the centers of mass of functional and structural DRCs. The distribution was obtained via a bootstrapping procedure. This gives a clear indication where functional connections occur in the absence of underlying structural connections. That is, where the center of mass for structural DRC does not coincide with the center of mass for functional DRC. For SF and SF-negRC networks (upper and lower row, respectively) the distances are comparatively low. For SF-RC networks (middle row) the distances are comparatively large (several degrees).

Results on empirical human connectivity

Employing the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) we examined the (estimated) entropy exhibited by a spin glass model whose underlying connec-

tivity matrix was a binary empirical structural connectivity matrix of the human brain (HUM) comprising 68 nodes derived from diffusion MRI data⁶. Additionally, the results were contrasted with those obtained for randomly rewired versions (HUM-RW) of the human connectivity matrix leaving the degree distribution intact but removing the rich club phenomenon as well as with artificial regular, random, and small world connectivity matrices of the same order and sparsity (Maslov & Sneppen, 2002). Figure 4.6 shows the entropy of the attractors for the set of networks comprising 68 nodes as a function of the global coupling strength for all classes of network architectures. Consistent with the simulations of network architectures comprising 24 nodes, the entropy obtained across a range of global coupling strengths was found to be higher for human structural connectivity as compared to small world, random, and regular network architectures. Furthermore, human structural connectivity (containing a rich club structure) revealed a higher level of entropy (mean entropy $\mu E = 8.64$) than rewired versions lacking a rich club ($\mu E = 7.97$) [$p < .001$ in a t-test with entropy at the critical coupling point ($G = 1$) as dependent variable].

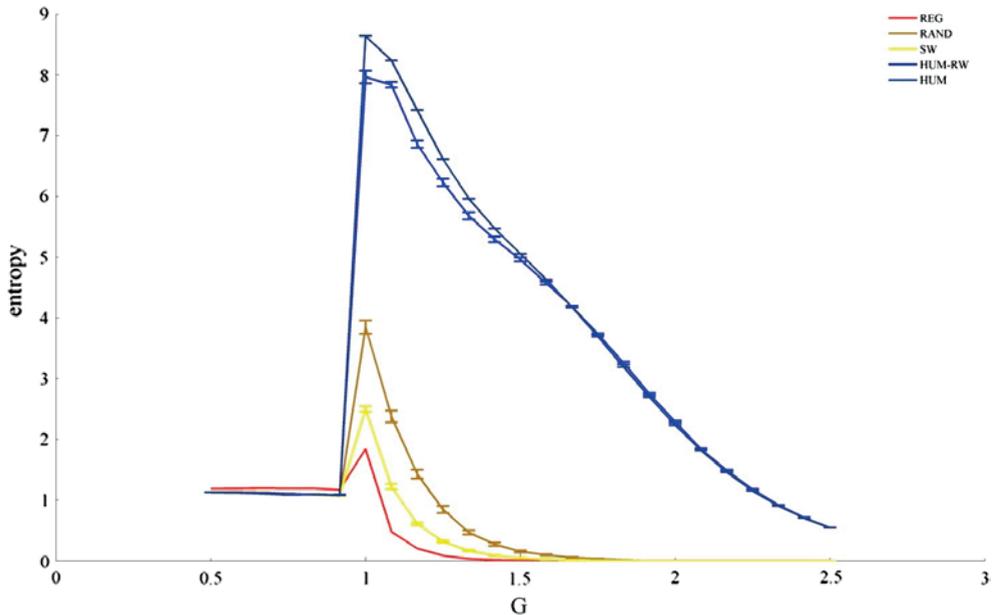


Figure 4.6: Entropy of the attractors of spin glass networks (68 nodes). The figure shows the entropy observed for each network class (including human anatomy) as a function of global coupling strength. Human rewired (HUM-RW, blue) network architectures which retain the hub nodes present in human anatomy show larger entropy levels over attractors than the corresponding small world (SW, yellow), regular (REG, red) and random (RAND, orange) networks. More importantly, however, the human (HUM, dark blue) network architecture which additionally features a rich club displays the largest levels of entropy.

⁶ We investigated the suitability of our implementation of the Metropolis-Hastings algorithm by obtaining the entropy of matrices comprising 30 nodes both analytically as well as using the algorithm. See figures A5.2 and A5.3 in the appendix.

We additionally examined the DRC obtained from empirical human structural and functional connectivity. We performed this analysis on functional connectivity as obtained from resting state fMRI. Figure 4.7 shows the empirical structural (a) and empirical functional DRC (b). As can be seen in figure 4.7b, functional clusters among low ($k = [4,10]$) degree nodes are prominent. Since these nodes together constitute 46% of all nodes it is their integration that leads to a diverse functional repertoire. While moderate-degree nodes share structural as well as functional connections (figure 4.6a and b), functional coupling among low-degree nodes must occur in the absence of direct structural connections. This observation overlapped with the simulated results of the artificial network architectures ($N = 24$ nodes, figure 4.5c).

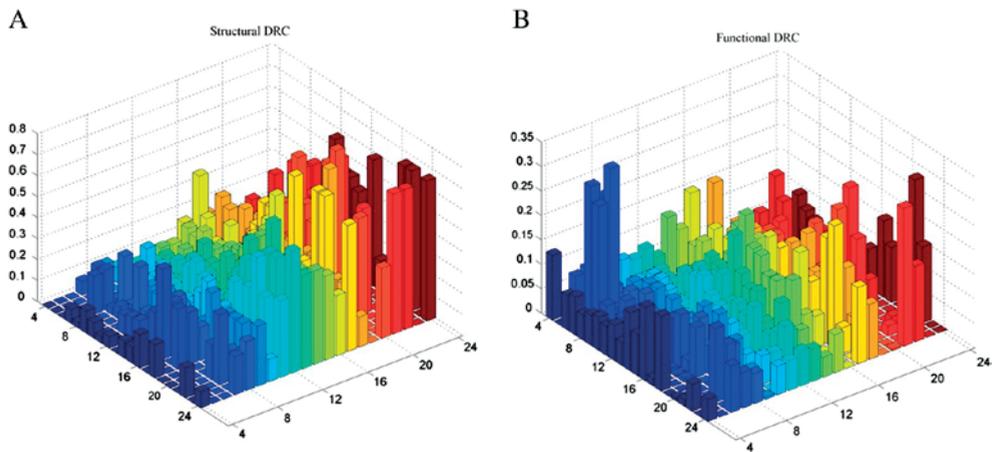


Figure 4.7: Degree-related connectivity in empirical human connectivity data. The figure displays degree related connectivity for empirically derived human structural and functional connectivity. Similar to the artificial networks (figure 5), the level of anatomical degree was divided into the categories low ($k = [4,10]$), moderate ($k = [11,17]$) and high ($k = [18,24]$). *Panel A*) The left panel shows human structural DRC, showing that moderate-to-moderate and moderate-to-high connections are most prominent. High-to-high connections exist, reflecting the rich club organization for the human connectome. *Panel B*) The right panel shows the functional DRC derived from empirical functional connectivity. Functional connections are present across all three categories with a strong cluster among low-degree nodes.

DISCUSSION

Our findings demonstrate that the presence of a central rich club core of mutually densely connected hub nodes constitutes a benefit to a network's functional repertoire. Across random, regular, small world, and three types of scale free networks as well as human empirical structural connectivity, the number of attractors (i.e. entropy) in the spin glass model was found to be largest for scale free networks displaying a rich club organization and hence higher than scale free networks that lacked dense connectivity between high degree hubs (i.e. SF and SF-nonRC, figure 4.3).

At the level of critical coupling, the spin glass model has been shown to exhibit a number of unique attractors representing the diversity of the system's functional repertoire (Deco et al., 2012). It has previously been suggested that a scale free architecture might be beneficial for a network with respect to obtaining a high level of entropy and thus a more diverse functional repertoire as compared to systems with an underlying regular, small world or random architecture (Deco et al., 2012). Extending these findings, our results now show that the presence of a central rich club within otherwise scale free architectures leads to an additional increase in the number of unique functional networks the system can sustain.

In the context of studies suggesting that higher numbers of attractors are linked to a more diverse repertoire of functional dynamics (Deco et al., 2012), our current computational findings may add to our understanding of rich club formation in neural systems. In the mammalian brain, and likely neural systems in general, the rich club has been suggested to form a high-capacity backbone allowing for dynamic routing of information (Collin et al., 2013; Harriger, van den Heuvel, & Sporns, 2012; Towilson, Vértes, Ahnert, Schafer, & Bullmore, 2013; van den Heuvel et al., 2012; Zamora-López, Zhou, & Kurths, 2011). In this context, we interpret our current findings as evidence for the presence of a structurally central rich club in the mammalian brain to have a positive effect on the diversity of the brain's functional repertoire and to permit functionally segregated cortical areas to become integrated when needed (Zamora-López et al., 2011). Indeed, our first examination of applying the spin glass model to simulate neural connectivity on the empirical anatomical connectivity data of the human brain suggests that the rich club may promote functional diversity specifically by supporting integration of information between low-degree nodes. As such, these simulation findings tend to suggest that a structural rich club may enable the dynamic formation of functional networks in which information can be integrated specifically for the performance of a given task (Bassett et al., 2012; van den Heuvel & Sporns, 2013a). Such an observation runs in parallel with modern theories of cognition hypothesizing the existence of a 'global workspace' in which segregated communities can exchange information through means of neural interaction (Baars, 2005). With regard to the shape of the functional repertoires, our results indeed suggest that the presence of a rich club enables the flexible coordination of low-degree nodes into unique functional networks. That is, the functional repertoire appears to feature a universally involved backbone in the form of almost constantly active rich nodes enabling and sustaining specific (re-)configurations among low-degree nodes. In doing so, rich nodes would naturally participate in many functional networks of the brain, which has indeed recently been suggested (de Reus & van den Heuvel, 2013; van den Heuvel & Sporns, 2013a). These findings are further in line with a previous study by Braga *et al.* (2013) which revealed that densely connected cortical areas carry functional traces of many functionally specific functional networks. This suggests that an anatomical rich club enables configurations of functional networks by sustaining activity of low degree nodes as well as by routing information among them. Our and related findings thus converge to the notion of a rich club to be beneficial for ensuring both

a high level of functional diversity as well as flexible integration and exchange of information among low-degree nodes (van den Heuvel & Sporns, 2013a).

Some points need to be considered when interpreting the findings of our study. First, the artificial network architectures studied semi-analytically are comparatively small, relating to the high computational cost of the investigation of the spin glass model in which every possible spin configuration is considered in detail (see appendix A5.1). However, the study of Deco *et al.* (2012) showed that while network size affects absolute levels of entropy, the relative effect that different network architectures have on entropy was shown not to be affected by size. Indeed, a similar effect was observed when comparing our $N = 24$ and $N = 30$ network results, as well as by the approximations of global entropy for human empirical data showing that the effect related to the rich club phenomenon is also present in biologically relevant networks of larger order.

Second, we note that our spin glass simulation results do not necessarily provide information on the level of flow of information between low degree and/or high degree nodes: the spin glass model as applied in our study can only convey information as to the activation state (i.e. 'up' or 'down') of nodes. That is, the functional DRC in our simulations only reflects in how far nodes of a certain degree range can form dynamic (re-)configurations of activation states and does not reveal in how far (re-)configurations of actual information flow occur. Our simulation results can therefore only to a certain extent be compared to empirical functional DRC observations derived from correlations between time-series. Future studies in this area might allow for more detailed investigation of temporal dynamics and might reveal whether activity fluctuations ride on top of a global high activation envelope (i.e. the up states), hence providing more information on the dynamical character of functional and effective connectivity. Nevertheless, co-activation is likely a necessary prerequisite for information flow between brain regions and the configurations of co-activation patterns provide a constraint on the patterns of such information flow. Additionally, our simulation results of a constantly active rich club and low functional DRC between rich club nodes are in agreement with recent empirical findings of high metabolic demand of high degree cortical areas (Bullmore & Sporns, 2012) and rich club nodes in particular and with observations of relatively low functional connectivity among rich club nodes as measured during resting state (Collin *et al.*, 2011). Rich club organization of brain networks has been suggested to potentially reflect the capacity of a neural system to facilitate neural interplay and neural communication among high degree hub regions when integration of information between low-degree segregated functional domains is needed (Braga *et al.*, 2013; de Reus & van den Heuvel, 2013; van den Heuvel & Sporns, 2013a, 2013b), a hypothesis consistent with our simulation results suggesting that a rich club can increase the overall capacity of the network to boost functional connectivity between low degree nodes. Future studies examining the effects of rich club organization in more detail using neural mass models which allow for the simulation of neural time-series and thus simulated functional couplings between regions would be of high interest to further elucidate the role of rich club organization to functional network organization and integration.

Third, the comparability of functional connectivity measured by mutual information in the spin glass model to the commonly used correlation metric for empirical resting-state fMRI measurements is debatable. Nonetheless, results stemming from both were highly consistent. In our paper we advocate that our findings indicate that functional integration of low-degree nodes is promoted by scale free systems that show rich club organization, including the (human) brain. However, our computational results should be interpreted largely as an attempt to provide a theoretical framework to be used for more in depth empirical as well as simulation research than we can provide here (Breakspear, Jirsa, & Deco, 2010; Friston & Dolan, 2010). A particularly interesting prediction of our results is that task related functional networks should be distinguishable mainly by specific co-activations of low degree cortical areas with the rich club being present and active across multiple task related functional networks. Another interesting topic of investigation for future studies would be the simulation of effects of disrupted anatomical rich club organization on the functional repertoire. Such studies might for example provide more insight into the functional effects of abnormal rich club formation as observed in patients with schizophrenia (Collin et al., 2011, 2013; Hoffman et al., 2005; van den Heuvel, Stam, Kahn, & Pol, 2009; Q. Yu et al., 2012).

In conclusion, our study provides computational support for the notion of rich club organization aiding the brain's overall repertoire of functional diversity (van den Heuvel & Sporns, 2013a, 2013b; Zamora-López et al., 2011). Networks with a scale free architecture combined with the presence of a central densely connected rich club revealed a higher level of entropy compared to networks with a random, regular, small world and indeed other scale free architectures. Additionally, our findings tend to suggest that networks featuring rich club organization allow for the functional integration of specialized cortical regions of low macroscopic degree. As such, our findings provide evidence in support of the notion that a rich club in neural systems may form a neural substrate that both enriches as well as modulates the brain's repertoire of distinct specialized brain functions to allow for the flexible integration of cortical areas into functional networks. Future research, investigating how additional graph theoretical properties affect the brain's functional repertoire, how they interact with a rich club, and how simulated and empirical disease related damage to the rich club might reduce the overall functional repertoire of a system, would be of particular interest.

ACKNOWLEDGEMENTS

Authors MS and RG were supported by the European Research Council under the European Union's Seventh Framework Programme (ERC-2010-AdG, ERC grant agreement n° 269853). Author GD was supported by the ERC Advanced Grant: DYSTRUCTURE. Author MPvdH was supported by a VENI grant of the Netherlands Organization for Scientific Research (NWO) and by a Fellowship of the Brain Center Rudolf Magnus.

Chapter

5

The Rich Club Can Organize State- Dependent Functional Network Formation by Engaging in Oscillatory Behavior

This chapter is an adapted version of the following article:

Senden, M., Reuter, N., Heuvel, M.P., Goebel, R.; Deco, G.
(in print). Cortical Rich Club Regions Can Organize State-
Dependent Functional Network Formation by Engaging in
Oscillatory Behavior. *NeuroImage*

ABSTRACT

Cognition is hypothesized to require the globally coordinated, functionally relevant integration of otherwise segregated information processing carried out by specialized brain regions. Studies of the macroscopic connectome as well as recent neuroimaging and neuromodeling research have suggested a densely connected collective of cortical hubs, termed the rich club, to provide a central workspace for such integration. In order for rich club regions to fulfill this role they must dispose of a dynamic mechanism by which they can actively shape networks of brain regions whose information processing needs to be integrated. A potential candidate for such a mechanism comes in the form of oscillations which might be employed to establish communication channels among relevant brain regions. We explore this possibility using an integrative approach combining whole-brain computational modeling with neuroimaging, wherein we investigate the local dynamics model brain regions need to exhibit in order to fit (dynamic) network behavior empirically observed for resting as well as a range of task states. We find that rich club regions largely exhibit oscillations during task performance but not during rest. Furthermore, oscillations exhibited by rich club regions can harmonize a set of asynchronous brain regions thus supporting functional coupling among them. These findings are in line with the hypothesis that the rich club can actively shape integration using oscillations.

INTRODUCTION

The human brain is characterized by a high degree of structural segregation allowing for designated information processing within specialized brain regions (Bear, Connors, & Paradiso, 2006; Flourens, 1842; Lashley, 1929). While this is beneficial for unimodal and automatic processing, higher cognition is hypothesized to require the globally coordinated integration of segregated brain regions into temporal functional networks (Baars, 2005; Deco, Jirsa, & McIntosh, 2011; Dehaene & Naccache, 2001; Ghosh, Rho, McIntosh, Kötter, & Jirsa, 2008). A high degree of integration is thus additionally required for efficient information processing (Damasio, 1989; Tononi, 2004). Studies of human, macaque, and other mammal cortices have shown the presence of a hierarchically higher module termed the rich club which is characterized by hubs with dense intra- and inter-modular connectivity (Colizza, Flammini, Serrano, & Vespignani, 2006; van den Heuvel & Sporns, 2011; Zamora-López, Zhou, & Kurths, 2009). The rich club forms a structural backbone mediating a majority of all anatomical paths between pairs of brain regions (van den Heuvel, Kahn, Goñi, & Sporns, 2012; Zamora-López et al., 2009), suggesting a prominent role for it in functional integration.

In accordance with this, recent functional magnetic resonance imaging (fMRI) studies have shown that cortical hubs carry traces of blood oxygen-level dependent (BOLD) activity of resting- and task related functional networks (Braga, Sharp, Leeson, Wise, & Leech, 2013; Leech, Braga, & Sharp, 2012), indicative of communication among brain regions being mediated by rich club regions. fMRI research has further shown that cortical hubs update their pattern of global functional connectivity in response to changing task demands (Cole et al., 2013). Additionally, simulation studies employing a steady-state attractor model have suggested that cortical hubs, and specifically the rich club, may allow the brain to sustain a large functional repertoire characterized by diverse configurations of peripheral, i.e. low degree, regions around a stable high-degree core (Deco, Senden, & Jirsa, 2012; Senden, Deco, de Reus, Goebel, & van den Heuvel, 2014). Similar observations have been made in studies using oscillatory neural mass models showing that the rich club can facilitate the synchronization among groups of cortical regions (Gollo, Zalesky, Hutchison, van den Heuvel, & Breakspear, 2015; Schmidt, LaFleur, de Reus, van den Berg, & van den Heuvel, 2015). The latter is especially interesting given that cortical regions can undergo transitions from asynchronous to oscillatory behavior as a result of interactions between cortical and thalamic neuronal populations in the presence of noise (Breakspear et al., 2006).

These findings suggest that rich club regions may utilize oscillations as a local control mechanism to organize brain regions into functional networks as they dynamically adjust their oscillatory behavior in response to changing task demands. Furthermore, if rich club regions are generally involved in functional integration, these adjustments should occur irrespective of cognitive domain. The aim of the present study is to investigate this possibility by combining fMRI measurements of resting and task states with simulations of a whole-brain model

whose cortical regions can exhibit local dynamics ranging from asynchronous fluctuations (noise diffusion) to structured oscillations. We restrict our investigations to oscillations falling within the infraslow frequency band (defined as the range from .01 Hz to .2 Hz; Vanhatalo et al., 2004) due to the slow nature of the BOLD signal. These oscillations might be a proxy for low-pass filtered oscillatory behavior at higher frequency bands. However, prior research suggests that the infraslow frequency band might also be relevant for functional network formation in its own right since oscillatory behavior in this band has been shown to modulate regions' excitability states and has been implicated in resting state and task execution (Hiltunen et al., 2014; Monto, Palva, Voipio, & Palva, 2008; Vanhatalo et al., 2004). To study the relevance of local oscillatory behavior for whole-brain dynamics related to cognition irrespective of a specific cognitive domain, we obtained fMRI data for a range of tasks, including an n-Back task (Kirchner, 1958), the Eriksen Flanker task (Eriksen & Eriksen, 1974), a mental rotation task (Shepard & Metzler, 1971), and a verbal odd-man-out task (Flowers & Robertson, 1985). These tasks were chosen to tap into working memory, executive function and inhibition, mental rotation, and semantic reasoning, respectively. These cognitive domains are conceptually different and their associated functional connectivity profiles have been shown to be minimally overlapping (Smith et al., 2009). The combined simulation and fMRI approach allows us to investigate the local dynamics of rich club regions as opposed to other cortical regions, whether rich club regions adapt their local dynamics in response to task performance as compared to rest, and how this relates to whole-brain functional coupling.

MATERIALS AND METHODS

Participants

Fourteen healthy subjects (8 females, age range = 22-43 years, mean age = 28.76 years) were recruited for resting state and task related fMRI measurements. All subjects had normal or corrected-to-normal visual acuity, were screened, and provided written informed consent prior to scanning.

Task and Stimulus Description

In the present study subjects underwent six functional runs. These runs consisted of a resting-state measurement, four individual task measurements including an n-Back (n=2) task (Kirchner, 1958), the Eriksen Flanker task (Eriksen & Eriksen, 1974), a mental rotation task (Shepard & Metzler, 1971), an odd-man-out task (Flowers & Robertson, 1985), and a task-switching paradigm wherein participants repeatedly performed each of the four tasks. Using the cognitive atlas (<http://www.cognitiveatlas.org>; Poldrack et al., 2011), tasks were selected based on how well they reflect a specific cognitive domain. The cognitive domains (working memory, executive function & inhibition, mental rotation, and semantic reasoning), in

turn, were chosen because they show distinctive patterns of mapping to the ten primary resting-state functional networks (Smith et al., 2009). All resting-state measurements preceded task-related measurements to prevent carry-over effects (Grigg & Grady, 2010). Resting-state runs lasted for 8 minutes during which subjects were instructed to close their eyes. The four individual task runs followed rest, lasted ~ 7 minutes each, and were counter-balanced across participants. The task-switching run lasted 9 minutes and was always performed last to allow participants to get familiar with performing each task in the scanner before being required to switch between them. Since this run was not the object of the present study it will not be discussed further. Descriptions of the remaining four tasks are as follows:

N-back Task. In a visual 2-back task subjects were presented with a sequence of abstract shape stimuli and instructed to indicate whether the currently presented stimulus matches the second to last stimulus in the sequence presented before. Abstract, snowflake-like, shapes were used in order to prevent subjects from translating a visual into a verbal representation and thus minimizing representational overlap with the verbal odd-man-out task. A total of 9 different snowflake-like shapes were shown in a quasi-random order. Shape presentation trials lasted for 1 second followed by a 1 second inter-trial interval. In total there were 192 trials forming a single task block preceded and followed by a 16 second rest period leading to a total run length of 416 seconds. Of the 192 trials 48 (25%) required a response indicating a match.

Flanker Task. In a modified version of the Eriksen Flanker task (Eriksen & Eriksen, 1974) subjects were presented with three rows of arrow stimuli pointing either to the right or the left. The center arrow was the target stimulus indicating whether a left or right response was required. The arrows surrounding the target stimulus were distractors whose direction was either congruent or incongruent with that of the target stimulus. Trials lasted for 1 second followed by a 1 second inter-trial interval. In total there were 192 trials forming a single task block preceded and followed by a 16 second rest period leading to a total run length of 416 seconds. Of the 192 trials 96 (50%) were incongruent thus requiring response inhibition.

3D Mental Rotation Task. In a mental rotation task subjects were required to mentally rotate a three-dimensional probe stimulus in order to confirm or deny an identity match with a target stimulus. Subjects could indicate a match between the two by pressing the 'right' button and a mismatch by pressing the 'left' button. Trials lasted for 3 seconds followed by a 1 second inter-trial interval. In total there were 96 trials forming a single task block preceded and followed by a 16 second rest period leading to a total run length of 416 seconds. Of the 96 trials 24 (25%) presented matching probe and target stimuli.

Odd-Man-Out Task. In a verbal odd-man-out task subjects were presented with three words and had to indicate which of the three semantically fits least with the others. The word being the odd-man-out in a specific trial was always either presented on the outer left or the outer right with the middle word forming a semantic anchor. Hence, the required response was a 'left' or 'right' button press, if the odd-man-out was presented on the left or the right of the anchor, respectively. Trials lasted for 3 seconds followed by a 1 second inter-trial interval. In total there were 96 trials forming a single task block preceded and followed by a 16 second rest period leading to a total run length of 416 seconds.

Stimulus Presentation

All tasks were programmed in Presentation® (Version 10; www.neurobs.com). Stimuli were presented to the subjects at a resolution of 1920x1200 pixels. All experiments were performed on a hardware configuration containing a Dell Optiplex 970 computer with a NVIDIA NVS 300 graphics card with OpenGL >2.0 support connected to a Panasonic PT EZ570E wuxga projector. The projected stimuli were first reflected off a mirror positioned behind the bore of the magnet before they were reflected off a second mirror located above the head coil. The projection pathway together with the scanner table positioning created a fixed maximum visual angle of 18.76° for all subjects.

Magnetic resonance imaging

Images were acquired at Maastricht Brain Imaging Centre (Maastricht University) on a 3T scanner (Tim Trio/upgraded to Prisma Fit, Siemens Healthcare, Germany). Anatomical data were collected prior to functional data with a T1-weighted MPRAGE imaging sequence (192 sagittal slices; Repetition Time [TR] = 2250 ms; Echo Time [TE] = 2.21 ms; Flip Angle [FA] = 9°; Field of View [FoV] = 256 x 256 mm²; 1 mm isotropic resolution. Functional images were acquired using a gradient-echo echo-planar imaging sequence (38 transversal slices; TR = 2000 ms; TE = 30 ms; FA = 77°; FoV = 216 x 216 mm²; voxel-size = 3 x 3 x 3.5 mm³; no slice gap; GRAPPA = 3).

Processing of (f)MRI data

Anatomical images were automatically processed with the longitudinal stream in FreeSurfer (Reuter et al., 2012; <http://surfer.nmr.mgh.harvard.edu/>) including probabilistic atlas based cortical parcellation according to the Desikan-Killany (DK) atlas (Desikan, Segonne et al. 2006). Initial preprocessing of each functional dataset was performed in BrainVoyager QX (v2.6; Brain Innovation, Maastricht, the Netherlands) and included slice scan time correction, (rigid body) motion correction, and high-pass filtering with a frequency cutoff of .01 Hz. Due to the use of preparation scans, none of the initial volumes needed to be discarded related to T1 equilibrium effects. Subsequently, functional images were registered to the subject's anatomical images. Using MATLAB (2013a, The MathWorks, Natick, MA), signals were then cleaned further by performing wavelet despiking (Patel & Bullmore, 2015) and regressing out a global noise

signal given by the first principal component of signals observed within the cerebrospinal fluid of the ventricles. Next, voxels were uniquely assigned to one of 68 cortical regions and an average BOLD signal for each region was computed as the mean time-series over all voxels of that region. Finally, since the initial and final 8 data points of all task runs constitute a rest-period, these data points were removed leaving 192 data points during which a task was continuously performed. For reasons of comparability, resting state BOLD signal was equally reduced to 192 data points. However, all analyses and simulations were repeated with the full resting state data set and did not produce different results.

Structural Connectivity

High-quality diffusion-weighted MRI data of 215 subjects was obtained from the human connectome project's (HCP) Q3 release (Glasser et al., 2013; Van Essen et al., 2012). White matter fibers were traced for each subject using generalized q-sampling imaging (GQI) and streamline tractography (Yeh, Wedeen, & Tseng, 2010) and the cortex was parcellated into 68 cortical regions based on the DK atlas (Desikan, Segonne et al. 2006). More details on these processing steps can be found elsewhere (de Reus & van den Heuvel, 2014). A weighted structural connectivity matrix was then generated by averaging streamlines over subjects, keeping only those entries which had positive values for at least 60% of subjects (de Reus & van den Heuvel, 2013), and resampling the data to follow a Gaussian distribution with a mean of .5 and a standard deviation of .15 (Honey et al., 2009). Gaussian resampling of, or alternatively log transforming, the data has recently been shown to enhance correspondence between diffusion tractography and in vivo animal tract-tracing measurements of anatomical connectivity (van den Heuvel et al., 2015).

Whole-Brain model

To examine the interplay of local and global dynamics during rest as well as in response to task demands we studied the behavior of interconnected regions represented by a neural mass model characterized by the normal form of a supercritical Hopf bifurcation and connected according to an anatomical structural connectivity matrix. We chose this description because it can capture transitions from asynchronous to oscillatory behavior. Neural mass models characterized by the normal form of a Hopf bifurcation had previously been shown to provide a good characterization of regional dynamics (Freyer et al., 2011; Freyer, Roberts, Ritter, & Breakspear, 2012). Briefly, the dynamic behavior of each region j was given by

$$\frac{dz_j}{dt} = z_j \left[a_j + i\omega_j - |z_j|^2 \right] + \beta \eta_j(t) \quad (5.1)$$

with z_j being complex and hence separable into a real (x_j) and imaginary (y_j) part

$$z_j = r_j e^{i\theta_j} = x_j + iy_j \quad (5.2)$$

and η being Gaussian white noise implemented as a Wiener process scaled by a factor $\beta = .02$. In equation 5.1 the bifurcation parameter a_j represents a local control variable which determined whether a region was primarily dominated by noisy fluctuations ($a_j < 0$) or by a stable limit cycle with frequency $f_j = \omega_j / 2\pi$ ($a_j > 0$). The bifurcation parameter is thus meaningfully interpretable as it gives an indication as to whether brain regions exhibit oscillatory behavior. Figure A6.1 in the appendix gives a detailed account of a single region thusly described. We also provide code for simulating the interaction among two mutually coupled regions online (<https://github.com/MSenden/Hopf.git>). Embedding these local dynamics into a large-scale model and separating real from imaginary parts of z leads to whole-brain dynamics defined by the following set of coupled equations:

$$\frac{dx_j}{dt} = [a_j - x_j^2 - y_j^2]x_j - \omega_j y_j + G \sum_i C_{ij} (x_i - x_j) + \beta \eta_j(t) \quad (5.3)$$

$$\frac{dy_j}{dt} = [a_j - x_j^2 - y_j^2]y_j + \omega_j x_j + G \sum_i C_{ij} (y_i - y_j) + \beta \eta_j(t) \quad (5.4)$$

In equations 5.3 and 5.4, the coupling strength G scaled the adjacency matrix C . Coupling strength is a global control variable separating global dynamics into a unistable regime in the form of a low activity ground state and a multistable regime. Optimal coupling generally occurs close to a bifurcation separating these regimes (Deco et al., 2011; Deco & Jirsa, 2012). Simulated BOLD signal of each area j was directly given by the real part of z (i.e. by x) with each region having its own characteristic frequency ω_j in the range from .04 Hz to .07 Hz (a subset of the infraslow frequency range) given by the averaged peak frequency of the empirical BOLD signals in that range (for a distribution of characteristic frequencies per task see figure A6.2 in the appendix). We chose this frequency range as it had previously been identified as reliably reflecting gray matter signals and being minimally affected by aliased physiological noise (Glerean, Salmi, Lahnakoski, Jääskeläinen, & Sams, 2012)

Optimization of Global Control Variables

In order to find the working point for which the model reproduces global dynamics exhibited by the cortex during rest, as well as in response to task demands, two parameters (control variables) needed to be optimized. The first was the previously mentioned global coupling strength G scaling the structural connectivity. The second global parameter was the bifurcation parameter A , a fixed value substituted for all local bifurcation parameters a_j . We characterized global cortex dynamics in terms of grand average functional connectivity (FC), dynamic functional connectivity (DFC), and metastability (MS). Optimization of global control variables

involved finding those values of G and A for which the model faithfully reproduced all three metrics as observed in our dataset (see figure 5.1a for a schematic overview).

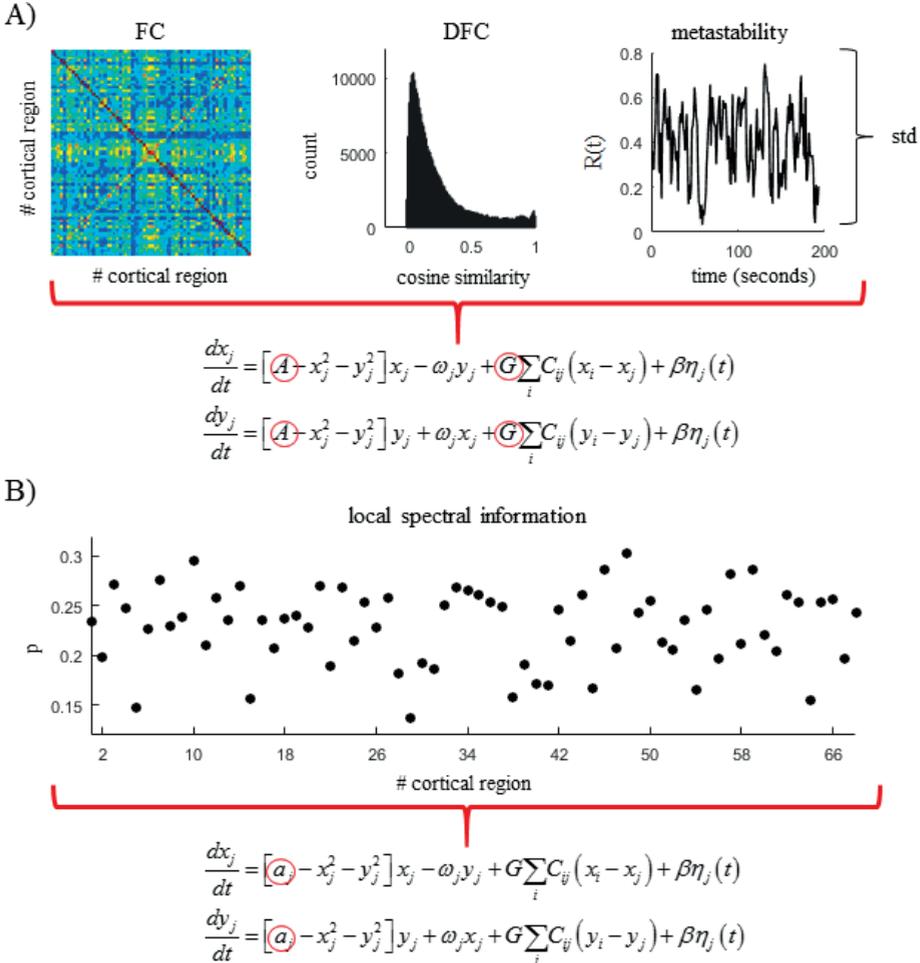


Figure 5.1: Schematic Overview of Parameter Optimization. *Panel A*) gives a brief overview of the optimization procedure for global parameter settings. Two parameters, coupling strength G and global bifurcation parameter A (both encircled in red), were adjusted in order for the model to reproduce empirically observed grand average functional connectivity (FC), the cumulative distribution of dynamic functional connectivity (DFC), and mean MS across subjects. Metastability is the standard deviation of the Kuramoto parameter $R(t)$ across time. *Panel B*) gives a brief overview of the optimization procedure of local bifurcation parameter a_j (encircled in red). After initializing all a_j to A (given the optimal working point A, G resulting from prior global optimization), values were adjusted according to a gradient descent strategy in order to capture the proportion p of power in a narrow frequency band with respect to a broad band observed for each cortical region j .

Grand average functional connectivity. This metric reflects the static component of the relationships between region-specific activation patterns. A grand average FC matrix per task (and rest) was obtained by calculating individual FC matrices in the form of pairwise Pearson correlation coefficients of bandpass-filtered (in the range from .04 Hz to .07Hz) regional BOLD signals for each subject, and subsequently averaging across individual subject's FCs. The same general logic applied to the grand average FC obtained for the simulated BOLD signal with the difference that rather than simulating separate subjects, simulation time corresponded to acquisition time multiplied by the number of subjects. The fit between empirical and simulated grand average FC was considered to be the Pearson correlation coefficient between the entries falling in the upper triangular of the respective matrices. Alternatively, model fit might be considered as the average of Pearson correlations between individual subject's FC and simulated grand average FC. The latter is worth noting since averaging over individual subjects' FCs can affect network characteristics (Moussa, Steen, Laurienti, & Hayasaka, 2012). However, both approaches lead to identical fit distributions across parameter space and we only report results from using the grand average FC.

Dynamic functional connectivity. This metric reflects the dynamics of functional couplings as short-lived global network states dissolve and may re-emerge at different moments in time (Hutchison et al., 2013). The DFC for each task (and rest) was given by the cumulative distribution of the pairwise similarity between instantaneous functional networks across subjects. Specifically, in an individual subject we first computed the analytic signal of each cortical region by applying the Hilbert transformation to its bandpass-filtered BOLD signal. This allowed for the calculation of instantaneous phases (i.e. phases observed at each moment in time) of cortical regions. Next, a phase difference matrix was obtained at each point in time by calculating the instantaneous phase differences between all pairs of cortical regions. These difference matrices were then transformed to similarity matrices by computing the cosine of their entries (figure 5.2a). A single phase similarity matrix reflects the functional connectivity among cortical regions observed at a single moment in time; i.e. the instantaneous functional connectivity (iFC). To estimate the similarity between functional connectivity observed at different moments in time, we calculated the cosine similarity of the upper triangular of iFC matrices between all pairs of time points (figure 5.2b). This results in a DFC matrix (figure 5.2c). The entries falling in the upper triangular of this matrix form the distribution of similarity among pairs of time points in terms of the functional connectivity observed at these moments in time (figure 5.2d). This procedure was repeated for all subjects with the final distribution of cosine similarity values being the aggregated distributions observed for individual subjects. We followed the same logic when calculating the DFC for simulated data. We calculated the Kolmogorov-Smirnoff distance between empirical and simulated distributions in order to evaluate their agreement.

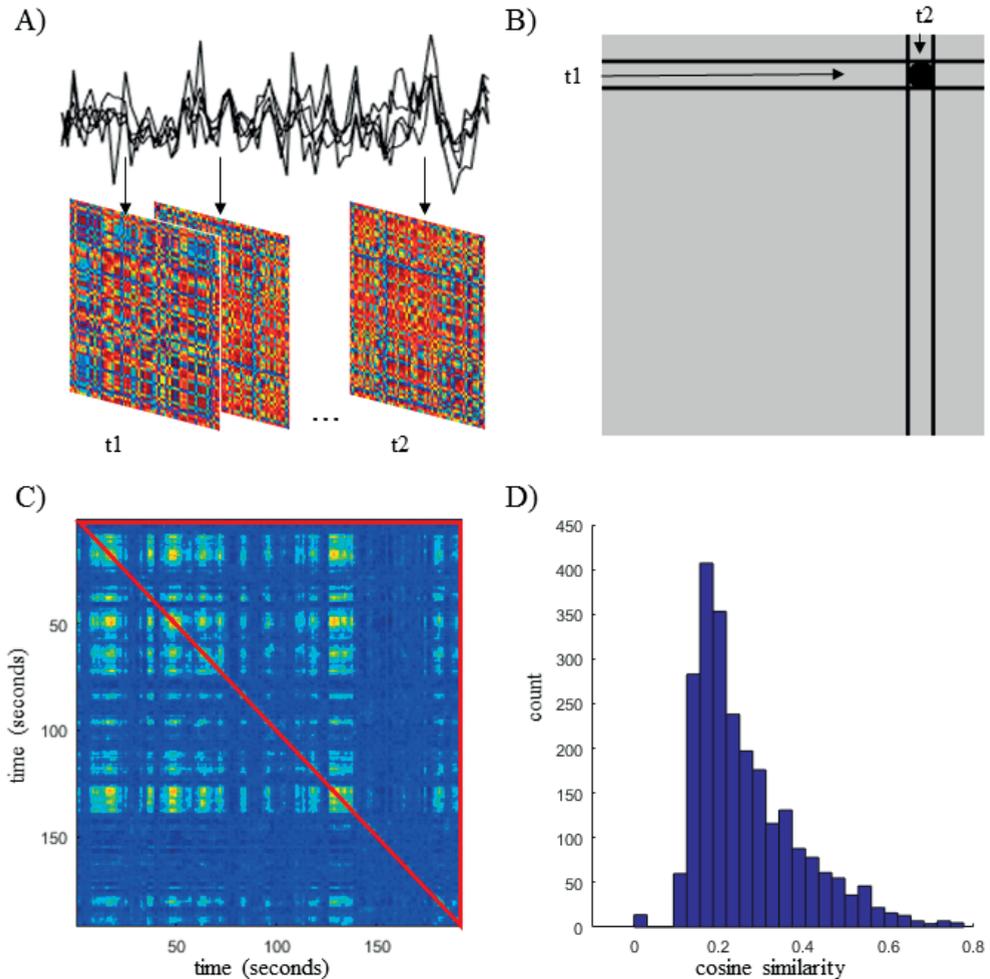


Figure 5.2: Schematic Overview of Dynamic Functional Connectivity Pipeline. *Panel A)* shows exemplary time signals as well as instantaneous functional connectivity (iFC) matrices at different moments in time. These iFC matrices reflect the cosine of instantaneous phase differences between all pairs of cortical regions at a given moment in time. *Panel B)* illustrates a timetime dynamic functional connectivity (DFC) matrix whose entry at t_1, t_2 is given by the cosine similarity between the upper triangular of iFC matrices observed at these moments in time. *Panel C)* depicts the full DFC matrix once the cosine similarity for all possible pairs of time points has been computed. *Panel D)* shows the distribution of cosine similarities observed in the upper triangular of the DFC matrix. The Kolmogorov-Smirnoff distance between empirically observed and simulated distributions indicates how well the model reproduces empirical DFC.

Metastability. The final metric used here reflects the overall variability of network states of the system (Wildie & Shanahan, 2012) and has previously been shown to be a good constraint for models whose nodes exhibit oscillatory behavior (Cabral, Hugues, & Deco, 2011; Cabral et al., 2014; Váša et al., 2015). Metastability in each task (and rest) was measured as the standard de-

viation of the Kuramoto order parameter observed over time. The Kuramoto order parameter $R(t)$ reflects the extent of synchronization exhibited among brain regions at a specific moment in time and is given by

$$R(t) = \left| \sum_{j=1}^n e^{i\varphi_j(t)} \right| / n \quad (5.5)$$

with $\varphi_j(t)$ being the instantaneous phase of each bandpass-filtered BOLD signal of regions j at time t and n the total number of brain regions. As for the estimation of DFC, instantaneous phases were calculated from analytic signals. For empirical data, MS was estimated for each subject separately and subsequently averaged. In the model it was obtained from the full-length signal. The difference between average empirical MS and model MS reflects how well the model reproduces this metric.

Optimization of the Local Control Variable

The region specific bifurcation parameter a_j was optimized by fitting spectral information of the empirical BOLD signals in each region (see figure 5.1b for a schematic overview). Specifically, local parameters were tuned to reproduce the region-specific proportions of power in a narrow band (0.04-0.07 Hz) with respect to a broad band (0.04-0.25 Hz). To do so, the power spectrum $P_j(f)$ for each node j in the narrow as well as in the broad band was calculated to obtain the proportion

$$p_j = \frac{\int_{.04}^{.07} P_j(f) df}{\int_{.04}^{.25} P_j(f) df} \quad (5.6)$$

Subsequently, the local bifurcation parameter was updated according to a gradient descent strategy

$$a_j = a_j + \eta(p_{j,empirical} - p_{j,simulated}) \quad (5.7)$$

A learning rate of $\eta = .1$ was used. Adjustment of this local parameter was only carried out after global parameters had been fit to assure that the procedure started with values in the vicinity of a global optimum. As for the optimization of global control variables, adjustment of the local bifurcation parameter was based on group data.

RESULTS

Rich Club

Rich club regions were identified from a binarized SC matrix obtained from setting all its non-zero entries to one. From this binary adjacency matrix rich club coefficients were calculated as the fraction of the number of existing connections between regions with degree larger than k to the possible number of connections among these regions (Colizza et al., 2006; Zhou & Mondragon, 2004). Next, the statistical significance of rich club coefficients for each degree k was determined by calculating the rich club coefficients for a set of 1000 degree-preserving rewired adjacency matrices (Maslov & Sneppen, 2002) and identifying the first k for which the rich club coefficient of the binarized SC was larger than the 95th percentile of the rich club coefficients corresponding to the rewired matrices. Candidate rich club regions were subsequently identified as those whose degree exceeded the first k level for which the rich club coefficient reached statistical significance. These candidate regions included the bilateral precuneus, the bilateral superior frontal cortex, the bilateral superior parietal cortex, and the right insula. To ensure that these regions were not only individually rich but also formed a dense club we calculated the internal density of this set of regions. Subsequently, we, in turn, removed each region and re-evaluated the internal density of the remaining set. If a region is part of the rich club its removal should hardly affect internal density. On the other hand, if a region is not part of the rich club its removal should lead to a sharp rise in density. Removal of the left precuneus, the right superior frontal cortex, left superior parietal, and right superior parietal each lead to decreases in internal density by 6.67 percent. Removal of the right precuneus and left superior frontal cortex each lead to a slight increase in internal density by 2.67 percent. These regions, therefore, form a dense club. However, removal of the right insula lead to an increase in internal density by 21.33 percent. Since ninety-five percent of density changes lay within ± 8.31 percent and density changes produced by removal of the right insula fell outside this interval, it was not considered a rich club region in this study. We validated this decision by removing the right insula from the SC matrix and observing no changes with regard to the rich club phenomenon.

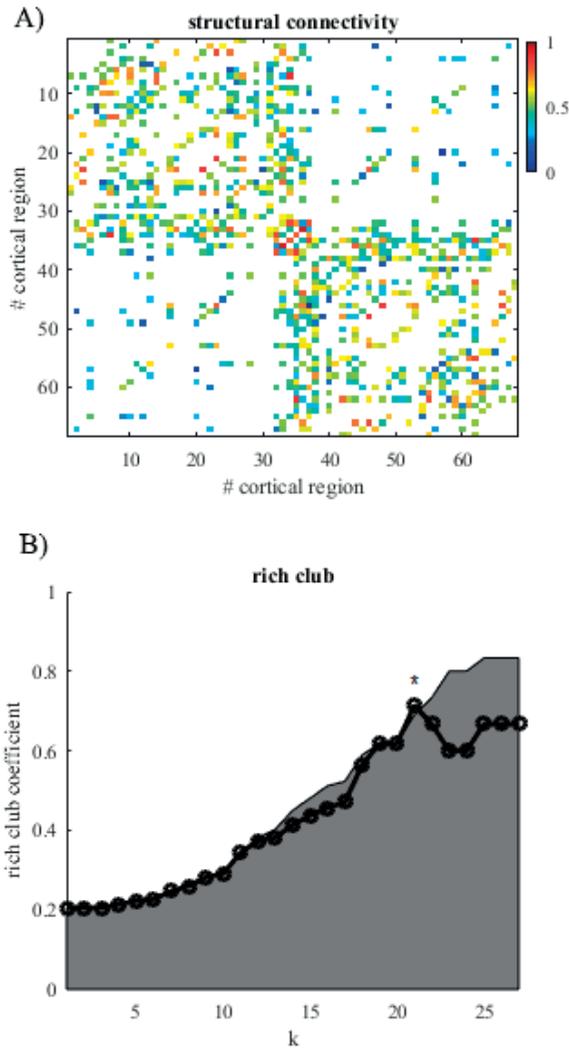


Figure 5.3: Structural data. *Panel A)* shows the weighted structural connectivity matrix consisting of 68 cortical regions. Regions are arranged according to the ordering given in the table in appendix 7 with the rich club in the center of the matrix. *Panel B)* shows the rich club coefficient of the structural connectivity matrix as a function of degree cutoff k (black line). The shaded region depicts the threshold the rich club coefficient needs to exceed to reach significance. Significance was reached for $k = 21$ (marked with an asterisk).

Whole-Brain Functional Dynamics and Model Fitting

Figure 5.4 shows the group level FC (A), DFC (B), and MS (C) observed for rest as well as for the four tasks which the model needed to reproduce. As can be appreciated from the figure, whole-brain FC patterns were highly similar. Indeed, all pairs correlated strongly with Pearson correlations in the range from .84 (rest with mental rotation) to .88 (flanker with mental

rotation). Dynamic FC representing the dissolution and re-assembly of short-lived functional connectivity patterns showed differences between resting and task states as well as among different task states. The distributions of cosine similarities (see methods for details) were significantly different according to a paired Kolmogorov-Smirnoff test for all pairs of rest to task as well as among tasks (P-values for all paired tests were $\ll .001$). Finally, metastability did not differ between resting and task states nor among different task states with the lowest P-value in a paired t-test equal to .26 (rest vs n-back).

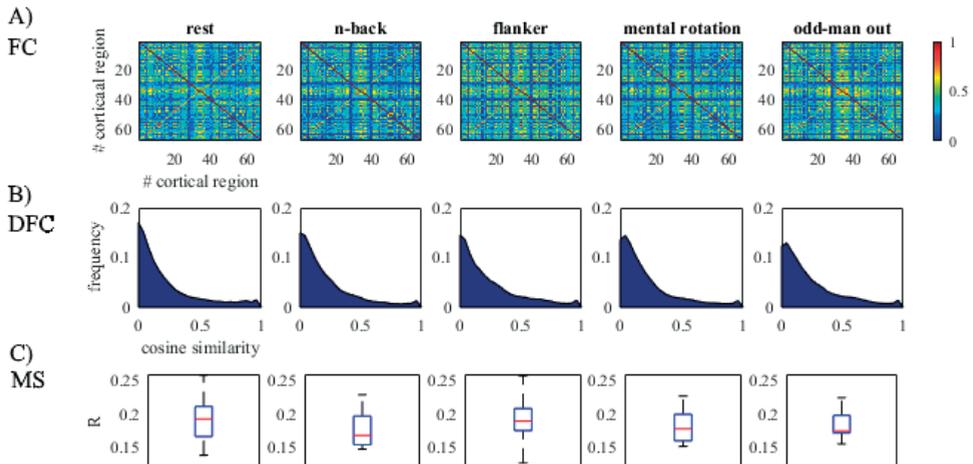


Figure 5.4: Empirical metrics. *Panel A)* shows the grand average FC for resting state as well as for the n-back, flanker, mental rotation, and odd-man out tasks. Regions are arranged according to the ordering given in table 1. *Panel B)* shows the distribution of cosine similarity as a metric of dynamic functional connectivity (DFC) for resting and all task states. Finally, *panel C)* shows boxplots of the standard deviation of the Kuramoto order parameter R as a metric of metastability observed in all functional states.

In order to find the optimal parameter settings for each task, we performed an exhaustive search of the parameter space. We examined global coupling G in the range from zero to .175 in 100 steps and global bifurcation parameter A in the range from minus to plus .5 in 30 steps. For all functional states, figure 5.5 shows the Pearson correlation between model and empirical FC (A), the Kolmogorov-Smirnoff distance between model and empirical DFC (B), and model MS (C) for the entire parameter space (i.e. all combinations of explored values for A and G). Figures A6.3-A6.5 in the appendix give a more detailed account of the three metrics obtained from simulations at different working points for resting state data. To obtain an overall measure of how well our simulations reproduce empirical data, we combined the three metrics. First we converted each metric to a distance between model and empirical data. For FC this involved subtracting the Pearson correlation from one. For MS this involved computing the difference between model and empirical MS values. The Kolmogorov-Smirnoff distance remained unchanged. Next we normalized each metric with respect to its range to adjust for

differences in scales. Finally, we averaged across these three normalized distance measures (figure 5.5d). Alternatively, the maximum across the normalized distance measures might be taken. Optimal parameter values correspond to the region where this global measure is minimized. We identified the same region using the average and the maximum to aggregate distance measures. This region was highly similar across tasks with a broad range of global coupling values but only a narrow band of bifurcation parameter values leading to good model fit. The latter was especially due to the Kolmogorov-Smirnoff distance between model and empirical DFC distributions as well as MS. In both resting and all task states, we found the optimal value for A to be just below zero implying that overall brain regions operated close to a transition from asynchronous to oscillatory dynamics (i.e., at the brink of a supercritical Hopf bifurcation) for all functional states.

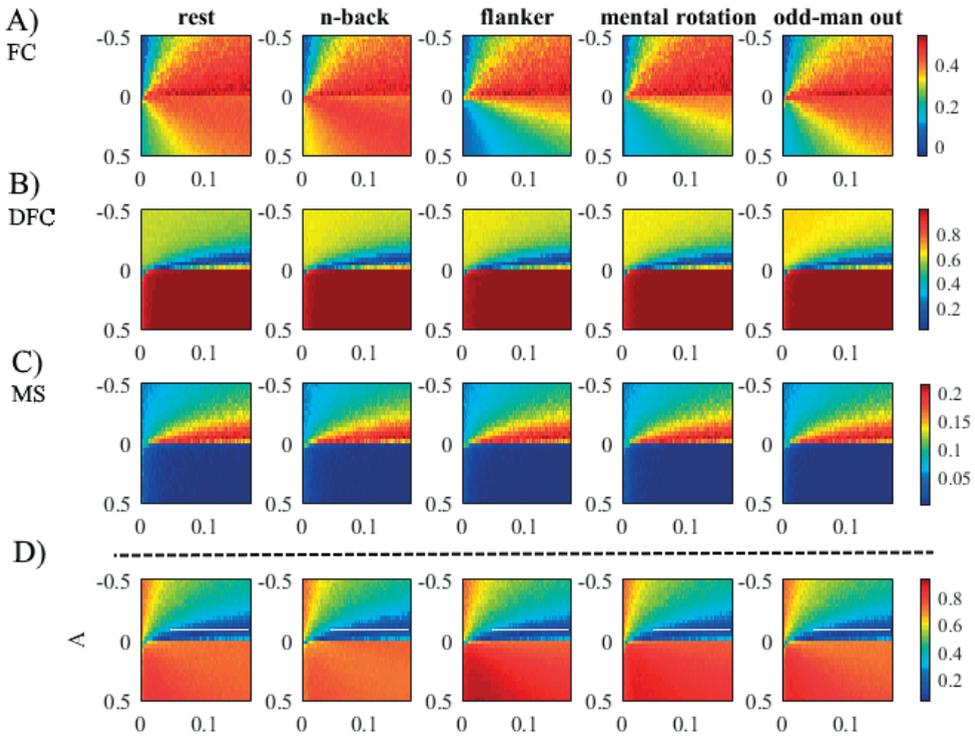


Figure 5.5: Parameter space exploration. This figure depicts the exploration of the parameter space defined by the global bifurcation parameter G and global coupling strength A for simulations of all functional states. *Panel A*) shows the correlation between empirical and simulation FC for all parameter pairings. *Panel B*) shows the Kolmogorov-Smirnoff distance between empirical and simulated DFC for all parameter combinations. *Panel C*) shows the MS observed for simulated functional states for all parameter pairings. Finally, *panel D*) shows an overall measure of model fit based on an aggregation of the three metrics expressed as a distance between model and empirical data. The range of parameter values for which our model simulations were performed is marked in white. We report results only for $G = .16$ and $A = -.0517$.

Local Dynamics

Next, we continued with the estimation of local bifurcation parameters a_j for all regions j within each functional state. Given the broad range of good global coupling values, we chose to estimate local bifurcation parameters for each task for a number of optimal global coupling values ranging from .04 to .16. However, since results were not qualitatively affected by this choice and since global coupling is conceptually a non-changing structural scaling factor, we present results only for the higher end of the range ($G = .16$). Local bifurcation parameter values were initialized to the average of optimal global bifurcation parameters values obtained across the range of optimal global coupling ($a_j = A_{opt} = -.0517$) thus ensuring that values were in the vicinity of a global optimum before starting the gradient descent. To evaluate the robustness of our local optimization procedure we also initialized a_j to random values in the range from -.5 to .5 (the results can be found in figure A6.6 in the appendix).

The results of local parameter estimations are shown in figure 5.6. For resting state, 60% of peripheral regions present with a bifurcation parameter above zero. As can be seen from the figure, rich club regions present with negative bifurcation parameters during rest. During task performance the number of peripheral regions presenting with a positive bifurcation parameter decreases. For the n-back, flanker, mental rotation, and odd-man out tasks 30%, 44%, 34%, and 44% of peripheral regions present with a positive bifurcation parameter, respectively. The mean bifurcation parameter across peripheral regions was $\bar{a}_p = -.03$ (95% CI [-.10, .04]) for rest, $\bar{a}_p = -.31$ (95% CI [-.41, -.20]) for the n-back task, $\bar{a}_p = -.13$ (95% CI [-.19, -.07]) for the flanker task, $\bar{a}_p = -.30$ (95% CI [-.41, -.20]) for the mental rotation task, and $\bar{a}_p = -.05$ (95% CI [-.10, -.01]) for the odd-man out task. While overall bifurcation parameter values drop for task states as compared to rest, rich club regions present with higher bifurcation parameters during task performance. The mean bifurcation parameter across rich club regions was $\bar{a}_{rc} = -.53$ (95% CI [-.59, -.48]), $\bar{a}_{rc} = .22$ (95% CI [.20, .25]), $\bar{a}_{rc} = -.05$ (95% CI [-.12, .01]), $\bar{a}_{rc} = .21$ (95% CI [.19, .23]), and $\bar{a}_{rc} = -.09$ (95% CI [-.12, -.06]) for rest and the four tasks, respectively. To further test whether rich club regions exhibit stronger positive changes in their bifurcation parameter than other cortical regions, we performed a cluster-based bootstrapping procedure. To that end we obtained the differences in bifurcation parameter values between task and rest for each region and computed the average difference across the set of rich club regions. Subsequently, we repeatedly (1000 times) sampled six regions from the entire set of cortical regions with replacement and calculated the average change within those sample. For all tasks the set of rich club regions exhibits changes far exceeding the 99th percentile of the thusly generated null-distribution. Specifically, the set of rich club regions presented with an average change of .75, .48, .74, and .45 for the n-back, flanker, mental rotation, and odd-man out tasks with their respective 99th percentiles being equal to .27, .28, .28, and .30. Figure A6.7 in the appendix shows a ranking of brain regions according to the changes they exhibit in local bifurcation parameters per task. Next, we evaluated in how far individual rich club regions presented with an increased bifurcation parameter during task performance as compared to rest. We

evaluated changes at the individual regions level by performing a blocked bootstrap test. The null-distribution for each region in each task was created from 1000 simulations of randomly generated rest and task samples and calculating the difference (task - rest) between local bifurcation parameters. Samples were created by first randomly drawing from the subject pool with replacement and then randomly placing one of each subject's two states in the *rest* and the other in the *task* sample repeatedly until each sample comprised 14 subjects. Local bifurcation parameters were then optimized in each of these samples. Figure 5.7 shows the results. For the n-back task all rich club regions exhibited a difference (rest-task) exceeding the 95th percentile of the bootstrap null-distribution. Furthermore, the difference observed for the left and right superior parietal cortices and the right superior frontal cortex exceeded the 99th percentile. For the flanker task differences observed in the left and right precuneus as well as the left and right superior parietal cortices exceeded the 95th percentile. Differences observed in the left and right precuneus exceeded the 99th percentile. For the mental rotation task differences in all but the left superior frontal cortex exceeded the 95th percentile and the right superior frontal cortex exceeded the 99th percentile. Finally, for the odd-man out task only differences observed for the left and right superior parietal cortices exceeded the 95th percentile while none exceeded the 99th percentile. The cluster analysis clearly indicates that rich club regions as a group exhibited more oscillatory behavior during task performance as compared to rest. However, the analyses of individual rich club regions shows that which rich club regions increased their oscillatory behavior and to what degree was task dependent.

Following up on these results we investigated whether cortex-wide functional connectivity was indeed related to oscillations. First, we examined the general effect of oscillations on simulated functional connectivity by calculating average global connectivity resulting from simulations in which all regions exhibited identical bifurcation parameters ranging from $-.25$ to $+.05$ with global coupling fixed at $G = .16$. Increasing the bifurcation parameter was accompanied by increased correlations between pairs of brain regions (see figure 5.8). Furthermore, positive bifurcation parameter values were associated with very strong correlations. Note that bifurcation parameter values just below zero; that is, at the optimal point determined previously, showed the largest range of correlations reflecting the vast dynamic range the system exhibited at this point. Subsequently, we identified for each functional state the subset of regions whose bifurcation parameter was estimated to be above zero as well as the subset of regions whose bifurcation parameter was estimated to be below zero, and calculated the average empirical functional connectivity within these two subsets. We obtained bifurcation parameters and functional connectivity separately by repeatedly (100 times) splitting the data in half and using one half of the data for the estimation of each. Figure 5.8b shows boxplots of the resulting estimates of average FC for the subsets of oscillating and non-oscillating regions for each functional state. In agreement with the finding that oscillations increased simulated functional connectivity values, average empirical FC was indeed higher among the subset of regions presenting with a positive bifurcation parameter for all functional states.

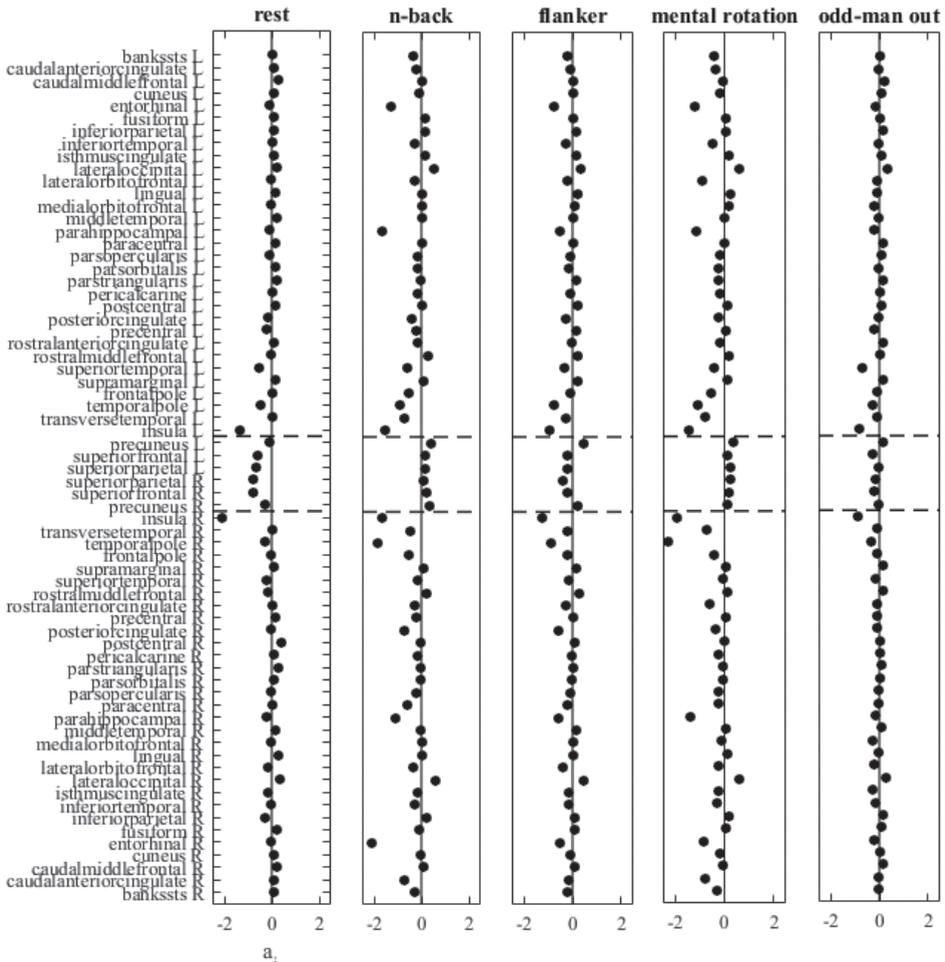


Figure 5.6: Local bifurcation parameters. This figure shows local bifurcation parameter a_i values observed for each of the 68 regions in the five different states. Regions are arranged according to the ordering given in the table in appendix 7 with rich club regions situated between the two red dashed lines. During rest, rich club regions display largely negative bifurcation parameter values whereas peripheral regions display positive and negative bifurcation parameters to a similar extent. During task performance, rich club regions largely display increased in their local bifurcation parameter values as compared to rest.

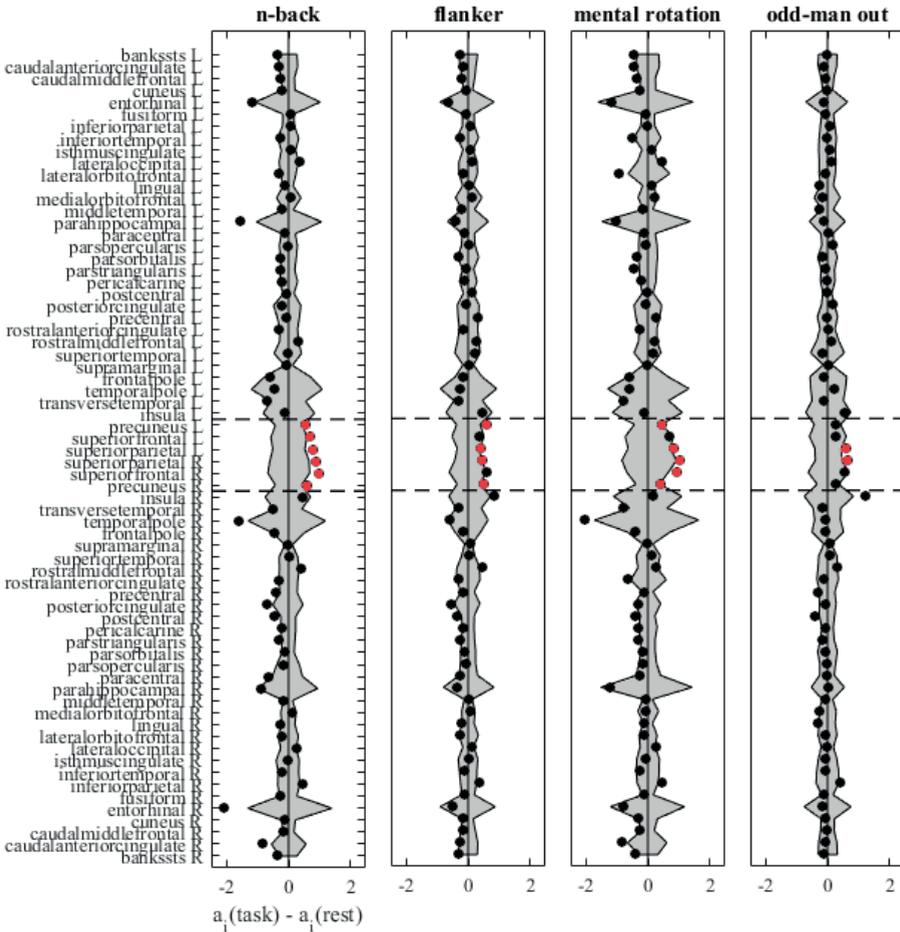


Figure 5.7: Differences in local bifurcation parameters values. This figure shows the change of regional bifurcation parameter values from rest for the four tasks $a_i(\text{task}) - a_i(\text{rest})$. Regions are again arranged according to the same ordering with rich club regions between the dashed lines. The shaded region depicts the null distribution of difference values as obtained from a blocked bootstrapping procedure. Rich club regions as a group display increases in their local bifurcation parameter values for tasks as compared to rest. Individual rich club regions exceeding the 95th percentile of a bootstrap null-distribution in expected differences between task and rest are marked in red.

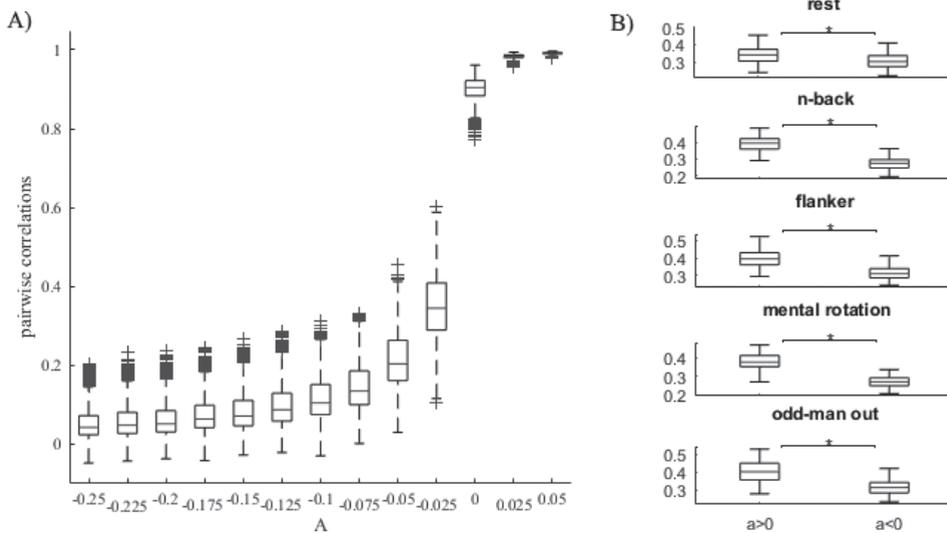


Figure 5.8: Relationship between bifurcation parameter and functional coupling. *Panel A)* shows the distributions of simulated functional connectivity (pairwise correlation) values for a range of global bifurcation parameter values A in the form of boxplots. A transition from negative to positive bifurcation parameter values is accompanied by overall higher pairwise correlation values, i.e. stronger overall coupling. The largest range of pairwise correlation values is observed for global bifurcation parameters just below zero. *Panel B)* shows boxplots of empirical FC values observed for regions presenting with a positive as compared to those presenting with a negative bifurcation parameter (a) value for each functional state. Larger FC values are observed for brain regions presenting with positive as compared to negative bifurcation parameter values, irrespective of the functional state.

Finally, we used the model to investigate whether oscillations exhibited by cortical rich club regions can be more effective in bringing about functional coupling than oscillations exhibited by peripheral regions alone. To this end we investigated in how far a set of brain regions exhibiting oscillatory behavior can synchronize cortical activation across the entire network. Two factors were of interest in this analysis: the first was the size of the set of brain regions exhibiting oscillatory behavior, whereas the second was whether this set included the rich club. The first factor was included since a sufficiently large number of peripheral (i.e. non rich club) regions exhibiting oscillations might be able to bring about synchronization without assistance of an oscillating rich club (as should be the case during rest). In terms of the first factor (size), a set comprising of 12, 18, 24, or 30 randomly selected peripheral regions were assigned a positive bifurcation parameter value ($a_{pos} = .5$) while the remaining regions were assigned a negative bifurcation parameter value ($a_{neg} = -.5$). This constituted at the same time the condition of only peripheral regions exhibiting oscillations. For the condition in which the rich club was among the set of oscillating regions, six randomly selected peripheral regions among the set of oscillating regions were replaced by rich club regions; i.e. the bifurcation parameter of these selected regions was set to $-.5$ while the bifurcation parameter of all rich club regions

was set to $+.5$. All peripheral regions had characteristic frequencies in the range from $.04$ Hz to $.07$ Hz whereas rich club regions were assigned a characteristic frequency at the center of this range ($.055$ Hz) in order to provide a pulse frequency. Each region's phase was initialized to zero but was effectively determined by the additive Gaussian noise term in equation 1. We measured synchrony among brain regions both in terms of phase and in terms of frequency. For synchrony in terms of phase we calculated the Kuramoto order parameter R (see methods for details) across brain regions at a frequency of $.055$ Hz. In terms of frequency we measured synchrony by calculating the power spectrum for each region in the aforementioned range normalized with respect to its maximum value. Subsequently, we averaged the power spectra over regions. If all brain regions exhibit fluctuations at the same frequency, the average power spectrum is characterized by a single sharp peak at the shared frequency ($\sim .055$ Hz). If, on the other hand, brain regions exhibit fluctuations at different frequencies, the average power spectrum is characterized by a broader peak. In order to characterize the breadth of the peak we fitted a Gaussian distribution to it and calculated its FWHM relative to the considered frequency range. We performed these simulations for 100 random selections of peripheral regions to exhibit oscillations nested in 100 randomly generated distributions of characteristic frequencies across regions. Hence, we performed 10,000 simulations for each combination of the number of oscillating peripheral regions and an oscillating or not oscillating rich club.

Figure 5.9 shows the relative FWHM as well as the Kuramoto order parameter R as a function of the total number of oscillating brain regions separately for the case that rich club regions exhibited oscillations and for the case that they did not. Synchronization both in terms of frequency as well as in terms of phase increased with a larger number of oscillating regions. Importantly, an oscillating rich club accelerated this effect. This effect was most pronounced for phase synchronization, but also apparent for frequency synchronization; especially if few peripheral regions were supplemented by the rich club. Such a configuration was characteristic of task performance. A larger number of oscillating peripheral regions (30 in this simulation) achieved good synchronization levels even in the absence of an oscillating rich club.

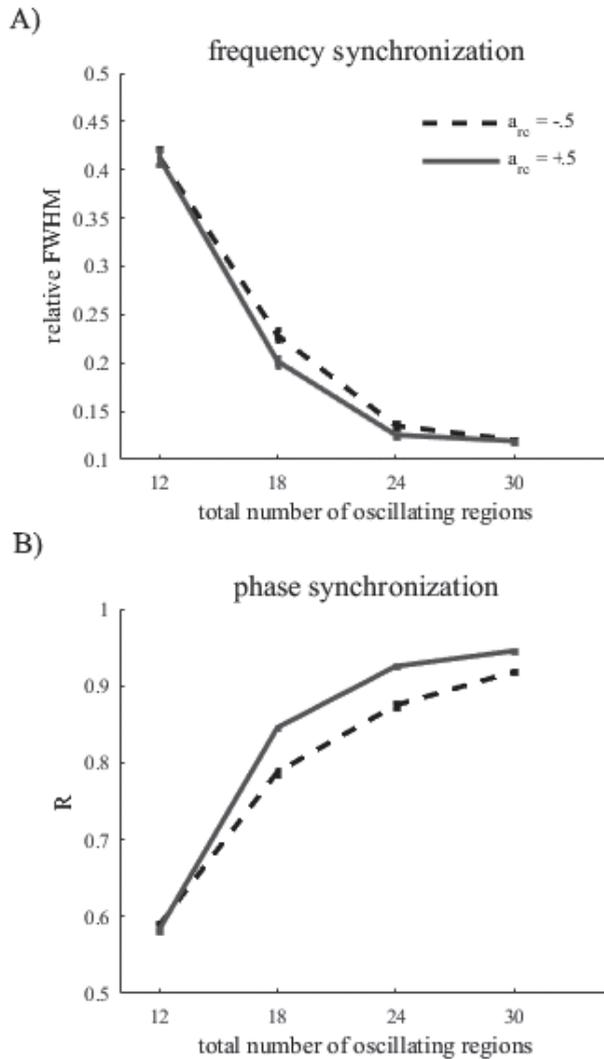


Figure 5.9: Synchronization. This figure depicts the extent to which brain regions are synchronized as a function of the number of brain regions set to have a positive bifurcation parameter value. Separate lines represent the cases where the set of regions having a positive bifurcation parameter value include the rich club (green) or do not include the rich club (red). *Panel A*) shows the extent of synchronization in terms of frequency as measured by the full width at half maximum of the power spectrum averaged across brain regions. Lower values indicate better synchronization. For the range of ~14 to 24 regions presenting with a positive bifurcation parameter value, it is beneficial if this set includes the rich club. *Panel B*) shows the extent of synchronization in terms of phases as measured by the Kuramoto order parameter R . Larger values indicate better synchronization. Unless a large number of regions is oscillating it is beneficial for synchronization to include the rich club among the oscillating regions.

DISCUSSION

Using a combined computational modeling and fMRI approach we investigated whether oscillations constitute a viable candidate mechanism through which cortical rich club regions can support functional network formation. Our results support this idea. First, brain regions whose local bifurcation parameter suggested that they exhibit oscillations showed an increased propensity to engage in functional coupling; both in terms of empirical as well as simulated functional connectivity. Second, in simulations, oscillations exhibited by the rich club provided a timing signal sufficient to synchronize brain regions operating at a range of idiosyncratic frequencies and phases. Finally, increased oscillatory behavior exhibited by rich club regions with simultaneous reductions in the number of peripheral regions engaging in oscillations demarcated empirically observed task states from resting state. In the present study we used tasks reflecting a wide range of cognitive functions; namely, working memory, executive function and inhibition, mental rotation, and semantic reasoning. Given that these cognitive functions are not only conceptually distinct but minimally overlapping with respect to their associated functional networks (Smith et al., 2009), it is reasonable to conclude that the observed changes in oscillatory behavior among the rich club as a whole are general rather than limited to specific cognitive processes. At the same time, differences in the degree to which individual rich club regions exhibit oscillations indicates that this behavior is adaptive to task demands. During task performance stable coupling among a sparse set of functionally relevant, specialized brain regions is required. During task performance, oscillating rich club regions might thus provide a pulse frequency to facilitate functional coupling among the comparatively few functionally relevant brain regions involved in task execution. Specifically, the small number of oscillating peripheral regions might be insufficient to sustain stable functional coupling unless supported by oscillations exhibited by rich club regions. This would be in line with previous simulation findings that the rich club can support functional coupling among peripheral brain regions (Gollob et al., 2015; Schmidt et al., 2015; Senden et al., 2014). During rest, on the other hand, rich club regions were not oscillating and could thus not impose a specific rhythm on the cortex. This behavior might be more conducive to flexible re-coupling among peripheral regions since it allows for larger groups of peripheral regions to engage in oscillatory behavior and hence to engage in (potentially short-lived) functional coupling without producing a fully coupled (epileptic) state. This might be advantageous for the exploration of the brain's functional repertoire (Deco & Jirsa, 2012).

A question at this point is whether the rich club is the origin of this coordination or provides the means for peripheral regions to exhibit control over the network. Previous research suggests the latter since the states of a complex system, and especially those which are difficult to reach, are controlled by regions of low degree rather than by hubs (Gu et al., 2015; Liu, Slotine, & Barabási, 2011). Nevertheless, the presence of hubs largely increases the general controllability of said system (Liu et al., 2011). Note that controllability of networks in general does not

necessarily require rich club organization but rather some densely interconnected structural backbone (e.g. a single hyperconnected node). In the brain this backbone is provided by the cortical rich club (van den Heuvel et al., 2012) whose role might be to enable peripheral regions to control the state, i.e. functional network configuration, of the cortex. As such, the rich club might provide a central workspace of information integration wherein peripheral brain regions compete for control of the system as recently proposed by Shanahan (2012). During rest, such conflict might be unresolved or, at least, recurring whereas task performance might necessitate a clear winner. This winner might then recruit the rich club to move the brain into the required functional state as well as to support communication among regions constituting this state. Our results suggest that this is dynamically achieved by moving the rich club into an (infraslow) oscillatory regime.

Our as well as previous findings (Gollo et al., 2015; Hiltunen et al., 2014; Monto et al., 2008; Schmidt et al., 2015; Vanhatalo et al., 2004) suggest that oscillations are relevant for global brain communication. However, it is currently unknown whether infraslow oscillations are immediately relevant for functional integration or are merely the filtered manifestation of a faster frequency band. It is principally conceivable, for instance, that stimuli presented at regular intervals impose an oscillatory activation pattern on task-relevant brain regions. While inter-trial intervals do not vary within tasks in our study, they differ between tasks. Furthermore, neither stimulation frequency aliases into the frequency band of interest (.04 Hz - 0.7 Hz). Finally, if the emergence of oscillations in our simulations were due purely to stimulation, we would expect largest changes in oscillatory behavior among task-specific and sensory regions rather than among rich club regions. These considerations do not rule out, however, that the observed infraslow oscillations constitute a proxy for oscillations occurring within alpha/beta or gamma frequency bands. This is a question that needs to be addressed using empirical data of higher temporal resolution. Due to the slow time scale of infraslow oscillations, it is at any rate unlikely that they carry the actual signals to be communicated between brain regions. A more likely function for them would be to align excitability states of brain regions (Hiltunen et al., 2014; Vanhatalo et al., 2004) and hence to establish communication channels among them. Such an interpretation for the potential role of infraslow oscillations would agree with recent electrophysiological findings which showed that the power envelope of higher frequencies, most prominently in the range from 8 Hz to 32 Hz, is itself modulated at infraslow frequencies (Hipp, Hawellek, Corbetta, Siegel, & Engel, 2012). These higher frequencies are far more likely to reflect the actual signal in line with a communication through coherence hypothesis (Fries, 2001; Fries, 2009; Michalareas et al., 2016). While our findings thus speak to the potential function of rich club regions to provide a timing signal for global brain communication, it remains to be investigated whether rich club regions also mediate and/or modulate high frequency signals exchanged among peripheral regions.

Furthermore, if infraslow oscillations are indeed meaningful, a question arises regarding the neural mechanism giving rise to them. The neural mass model employed here gives a

descriptive rather than a mechanistic account of local dynamics. This provides the appropriate level of detail for investigating the question of which local dynamics are exhibited by interconnected brain regions and how these change from resting to task states in an fMRI context. However, future work is needed to address the neural mechanism underlying these dynamics. A candidate mechanism comes in the form of interactions between cortical and thalamic neuronal populations (Breakspear et al., 2006; Freyer et al., 2011, 2012). A neural mass model implementing these interactions in the presence of state-dependent noise has recently been shown to account for known spontaneous transitions between two distinct modes of power in the alpha frequency band (Freyer et al., 2011; Freyer, Aquino, Robinson, Ritter, & Breakspear, 2009). Given the relationship between infraslow oscillations and the alpha and beta rhythms (Hipp et al., 2012), the slow dynamics examined here might reflect changes in the power envelope resulting from these transitions. Finally, since relay nuclei in the thalamus can themselves exhibit infraslow oscillations (Lőrincz, Geall, Bao, Crunelli, & Hughes, 2009), a straightforward extension of the cortico-thalamic neural mass model may allow it to account for both spontaneous as well as structured, i.e. infraslow oscillatory, transitions between low and high power modes in the alpha and/or beta bands.

Apart from the above, our results lead to a range of further questions and predictions. First, if the rich club is indeed a locus of competition among peripheral regions for network control, then it should receive a high degree of input but only provide sparse output; namely output reflecting communication among the winning set of regions. Furthermore, this should be task dependent both in terms of the total amount of input and output each rich club region receives and sends as well as in terms of the exact patterns of in and out relations. Second, if coupling via oscillations reflects the establishment of task-relevant communication channels, then one should expect to find information transfer among these regions (possibly mediated through the rich club) coded at higher frequencies, and that the informational content of these signals is relevant for, and reflects, the task at hand. Related to this, the information content as well as the degree to which rich club regions exhibit oscillatory behavior might also reflect task difficulty, a factor we did not take into account here. Finally, an important question future research should address is how local and whole-brain dynamics vary across individuals and how this relates to individual differences in the underlying structural connectivity. This would call for fMRI and diffusion-weighted MRI data to be acquired in the same subjects. Additionally, a larger sample size would be helpful in characterizing this variability. Furthermore, fitting model parameters to individual subject data would require longer runs thus putting practical limits on the number of tasks for which data can be acquired.

In conclusion, we find that (infraslow) oscillations constitute a potential mechanism for rich club regions to shape whole-brain functional coupling in a functionally specific manner. This lends further support to the possibility that rich club regions are crucial for the dynamic integration of otherwise highly segregated and specialized information processing and hence to a conceptualization of the rich club as the structural underpinning of a global neuronal

workspace (Dehaene & Changeux, 2011; Harriger, van den Heuvel, & Sporns, 2012; Shanahan, 2012). Furthermore, our results emphasize the importance of the integration of large-scale computational models with locally specific dynamics in conjunction with empirical research for investigating whole-brain functional network states. While all empirical metrics could be reproduced by our model using identical global parameters for each task, fitting and interpretation of parameters controlling local dynamics enabled us to demarcate resting and task states.

ACKNOWLEDGEMENTS

Authors MS and RG were supported by the European Research Council under the European Union's Seventh Framework Programme (ERC-2010-AdG, ERC grant agreement no. 269853). Author GD was supported by the ERC Advanced Grant: DYSTRUCTURE(n. 295129), by the Spanish Research Project SAF2010-16085 and by European Community's Seventh Framework Programme under the project "BrainScales" (project number 269921). Author MPvdH was supported by a VENI grant of The Netherlands Organization for Scientific Research (NWO) (451-12-001) and by a Fellowship of the Brain Center Rudolf Magnus. The authors declare no competing financial interests.

General Discussion

Chapter 9

Different abstractions from the same wholes capture different aspects of the reality but also leave us with different blindnesses.

Richard Levins

The objective of this thesis was to present interdisciplinary research guided by theory originating from computational neuroscience and constrained by observations obtained from fMRI. Within this larger objective, the first part of the thesis was dedicated to theory testing in the context of the neuroscience of vision while the second part was dedicated to theory building in order to further our understanding of the relationship between structure and function. In what follows the results of each part are summarized. Subsequently, I will examine whether the goals set forth in the introduction were met, relate the approaches presented in this thesis to computational neuroimaging, and discuss potential future research directions.

VISION

The work reported in the first part was concerned with developing a framework suited for evaluating the detailed spatial activation profile predicted by computational models using neuroimaging data. Since the approach detailed in this part requires population receptive fields (pRFs; Dumoulin & Wandell, 2008) as a bridging principle between simulated and empirical data, chapter 2 was dedicated to quality assurance with regard to their estimates. This includes not only robust estimation of pRFs but also their temporal consistency. The accuracy with which pRFs can be estimated in a functional magnetic resonance imaging (fMRI) setting requires an accurate estimation framework. Therefore, adequate decisions with regard to stimulus choice and mode of presentation were examined. Additionally, it was evaluated whether the stimulation protocol should entail mean luminance periods and whether it is advantageous to average the blood oxygenation level dependent (BOLD) signal across stimulus cycles. By systematically studying the effects of these decisions on pRF estimates in an empirical as well as simulation setting, the research presented in chapter 2 revealed that a bar stimulus presented at random positions and interspersed with mean luminance periods is generally most favorable among the examined stimuli. Chapter 2 furthermore confirmed the assumption of temporal consistency of pRF estimates by showing that parameters fit in two temporally separated sessions were highly similar.

Chapter 3 demonstrated that fMRI can be used to compare the detailed spatial activation profile of representations predicted by competing computational models if empirical and simulated data are represented within the same reference space. It builds on the previously proposed common brain space (CBS) which is an idealized, retinotopically organized, surface representation of visual cortex. Using pRF estimates, the functional correspondence between voxels and model units in the CBS was established at a resolution sufficient to contrast spatial activation profiles of different computational models. The approach was then exemplified by contrasting two models accounting for simultaneous brightness contrast (Blakeslee et al., 2015; Blakeslee & McCourt, 1999, 2004; Hurlbert & Poggio, 1988; Rudd & Arrington, 2001; Rudd & Popa, 2007; Rudd, 2001). The first model accounts for simultaneous brightness contrast in

terms of spatial filtering (Blakeslee et al., 2015; Blakeslee & McCourt, 1999, 2004). The second model proposes an edge integration mechanism for the effect wherein the representation of surface luminance is modulated by lateral spread of edge information (Peters et al., 2010). In the study presented in chapter 3, each model was supported in one of two conditions of the simultaneous brightness contrast but not both. Specifically, fMRI activation in response to a gray region surrounded by white background supported the EDGE model whereas fMRI activation in response to a gray region surrounded by black background supported the ODOG model. These results suggest that neither model is individually sufficient to account for the effect. The fact that the performance of the two models was complementary with respect to the two conditions does suggest, however, that the mechanism implemented in each has some merit. This points to the possibility that early visual cortex employs both mechanisms to some degree and that a carefully designed hybrid model might be able to account for fMRI activity observed in both conditions. The common brain space approach was thus conducive to demonstrating the limitations and strengths of each model and additionally pointed to their complementary nature. To conclude, the results presented in the first part of this thesis show that population receptive fields provide a robust basis for functional alignment of empirical and simulated data within a common brain space and that models can be meaningfully compared within the space.

STRUCTURE-FUNCTION

The studies presented in the second part of this thesis were aimed at elucidating the relationship between the brain's structure and function. They were specifically concerned with the role of the cortical rich club in the globally coordinated integration of segregated brain regions into temporal functional networks. Chapter 4 used a steady-state neural network model to examine the effects of rich club organization on the functional repertoire of physical systems in general, and the brain in particular. Using simulations, it was shown that the presence of a rich club increases the diversity of the functional repertoire over and above the effects produced by scale free topology alone. Within the networks' overall functional repertoire rich nodes were shown to be important for enabling a high level of dynamic integration of low-degree nodes to form functional networks. These results suggested that the rich club serves as an important backbone for numerous co-activation patterns among peripheral nodes of the network. In addition, applying the model to empirical anatomical data of the human brain, it was shown that the positive effects on the functional repertoire attributed to the rich club phenomenon can be observed for the brain's structural connectivity as well. These results were in line with a computational theory proposing that the rich club provides a central workspace for the integration of otherwise distributed information processing (Har-riger et al., 2012; Shanahan, 2012).

In order for the rich club to fulfill this role it must dispose of a dynamic mechanism by which it can actively shape networks of brain regions whose information processing needs to be integrated. Chapter 5 examines a potential candidate for such a mechanism coming in the form of oscillations which might be employed to establish communication channels among relevant brain regions. This possibility was explored using an integrative approach combining whole-brain computational modeling with neuroimaging. Specifically, the study presented in chapter 5 investigated the local dynamics model brain regions need to exhibit in order to reproduce (dynamic) network behavior empirically observed for resting as well as a range of task states. It was found that rich club regions generally exhibit oscillations during task performance but not during rest. This demarcation suggests that oscillations exhibited by rich club regions are functionally relevant. Furthermore, simulations revealed that an oscillating rich club is sufficient to harmonize a set of asynchronous brain regions thus supporting functional coupling among them. These findings were in line with the hypothesis that the rich club can actively shape integration using oscillations. The results presented in the second part of this thesis in conjunction with converging results from connectomics (de Reus & Heuvel, 2013; Harriger et al., 2012; van den Heuvel et al., 2012; Zamora-López et al., 2011), neuroimaging (Braga et al., 2013; Leech et al., 2012), and computational neuroscience (Gollo et al., 2015; Gu et al., 2015; Liu et al., 2011; Schmidt et al., 2015) suggest that the rich club might be the structural underpinning and oscillations the dynamic mechanism of what might at a cognitive level be considered to be a global workspace (Baars, 1993, 1997, 2002, 2005; Dehaene & Changeux, 2011).

WERE THE GOALS OUTLINED IN THE GENERAL INTRODUCTION MET?

The goal of the present thesis was to present work exemplifying that by integrating fMRI with computational modeling both disciplines benefit from the resulting theory driven research. The first part of the thesis contrasted two models accounting for surface perception by postulating a spatial filtering and an edge integration mechanism, respectively. Models implementing both mechanisms have been widely tested using psychophysical experiments (Blakeslee et al., 2015; Blakeslee & McCourt, 1999, 2004; Hurlbert & Poggio, 1988; Rudd & Arrington, 2001; Rudd & Popa, 2007; Rudd, 2001) to the extent that any attempt at further testing, or indeed refining, these models using psychophysics would yield diminished returns (Levins, 2006). Extending the methodological domain by incorporating ultra-high field fMRI held the promise of shedding new light on the issue (Chatham & Badre, 2014; Greenwald, 2012). Since the two mechanisms differ with respect to the algorithms they propose but not with respect to which information they represent (the two aspects of Marr's second level; Marr, 1982), the decoding abilities of standard fMRI were not useful for distinguishing among them. Rather, ultra-high field fMRI was needed to be sensitive to different implications following from the specific

implementations of each mechanism with regard to how information is spatially represented. Chapter 3 showed that the increased spatial resolution of ultra-high field fMRI was sufficient to discriminate between spatial profiles predicted by two model implementations, showing that this approach is viable. Computational modeling thus benefitted from the interdisciplinary work presented in chapter 3 of this thesis. At the same time, fMRI research benefitted in a general sense by being supplied with two potential mechanisms accounting for surface perception along the lines of which empirical results could be interpreted.

The second part of the thesis was concerned with building a theory regarding the functional role of the rich club for whole-brain dynamics (functional integration; chapter 4) and the general strategy through which it fulfils this role (oscillations; chapter 5). The second aspect was in part investigated through observing the dynamics exhibited by rich club regions during a range of whole-brain functional states. For this both coverage and localization capacity provided by fMRI was essential. However, none of three commonly employed univariate fMRI measures were sufficient to even demarcate functional states and while a multivariate pattern across the three measures might have been sufficient for demarcation, that in itself would not have revealed how rich club regions contribute to this demarcation. Only by implementing a specific candidate mechanism in a large-scale brain model could this question be addressed. It was thus not merely beneficial but essential to supplement fMRI research with computational modeling. At the same time, while a pure modeling approach could have shown that the proposed mechanism might indeed constitute a viable strategy through which the rich club fulfils its integrative role, empirical data was required to provide evidence that the brain indeed utilizes the mechanism. Therefore, computational modeling, too, benefitted from the interdisciplinary work presented in chapter 5.

The latter example highlights the obverse of the danger of “convenience experimentation” (Krohs, 2012) within neuroimaging: the temptation to conduct “convenience modeling” within computational neuroscience. Specifically, it might be tempting to treat a model as the object of interest rather than a representation thereof. Getting carried away by abstractions of the brain one might end up learning much about the abstraction and little about the brain itself (Levins, 2006). Thus, while studying a model can be highly informative, it is important to translate one’s findings back to the brain either by constraining the model with or testing its predictions against empirical results. With regard to the present thesis, it was thus equally important to provide empirical results in the form of fMRI data for computational models as it was to provide theoretical guidance in the form of computational models for fMRI research.

In conclusion, the work detailed in the present thesis was indeed successful in combining fMRI with computational neuroscience in order to conduct theory-driven research. Both disciplines, furthermore, benefitted from their integration. As a result the reported work presents two additional approaches towards interdisciplinary research and can thus complement existing approaches such as computational neuroimaging.

RELATION TO COMPUTATIONAL NEUROIMAGING

Computational neuroimaging refers to the development of quantitative models that predict fMRI activity from a (visual) stimulus (Wandell & Winawer, 2015; Wandell, 1999). The employed models are input-referred and allow for mapping the response profiles of cortical regions in terms of model parameters characterizing the input (Wandell & Winawer, 2015). For example, the parameters of population receptive field models, the most prominent tool within computational neuroimaging, characterize the input (sensory) space to which a neuronal population is responsive. The strength of this approach is that it allows for the formulation of empirical results in terms of model parameters and can thus be used to compare results stemming from different instruments, sessions, or groups (e.g. clinical; Binda et al., 2013; Dilks et al., 2009; Haak et al., 2014).

In the first part of this thesis, this strength was exploited to establish correspondence between computational models, whose response profiles were known, and brain regions, whose response profiles were mapped. A crucial step in the model comparison approach proposed in the first part of the thesis thus involves computational neuroimaging. The research presented in chapter 5 also bears some resemblance to computational neuroimaging in that it is reminiscent of the mapping procedure. Specifically, in both approaches a set of parameters is adjusted to minimize the distance between fMRI activity (measures) observed empirically and predicted by the model given a specific set of parameter values.

There is an important distinction between computational neuroimaging and the work presented here, however. The models employed differ in terms of their semantics; that is, with respect to the target they represent (Frigg & Hartmann, 2012). Input-referred models represent input space and can provide information on what aspects of this space are represented by different parts of the brain. These models do not propose any mechanistic theory as to how these representations come about. The models discussed in this thesis, on the other hand, do represent mechanistic theories, formulated either in terms of their specific algorithmic implementations (first part) or in terms of their general logic (second part).

FUTURE DIRECTIONS

As the last chapter of this thesis draws to a close, it is fitting to emphasize the cyclical nature of science and end with new beginnings. Given what has been presented in the present thesis a number of research avenues present themselves which either concretely extend research reported here or are generally in spirit with the interdisciplinary approach. With regard to the former, a particularly alluring project related to the first part of the thesis would be to use the common brain space in conjunction with an in-depth examination of anatomical and physiological properties of early visual cortex in order to build and evaluate an array of com-

putational models implementing a hybrid edge-integration-plus-spatial-filtering mechanism for surface perception. In a similar vein, it would be interesting to use the common brain space to build and/or evaluate models of border ownership (Craft, Schütze, Niebur, & Heydt, 2007; Jehee, Lamme, & Roelfsema, 2007) and figure-ground segregation (Roelfsema, Lamme, Spekreijse, & Bosch, 2002) as they make predictions with regard to spatial activation profiles in early visual cortex. A final and very important direction for future research related to the common brain space and computational model comparison using fMRI data in general, concerns critically evaluating and improving models employed to bridge from (simulated) neuronal to fMRI signals including models of population receptive fields (Kay et al., 2013; Zuiderbaan, Harvey, & Dumoulin, 2012) and of the (spatio-temporal) hemodynamic response (Havlicek et al., 2015; Logothetis et al., 2001; Shmuel et al., 2007).

With regard to the second part of this thesis, future research needs to address at least three open questions. The first is concerned with the relationship between slow and fast oscillations potentially exhibited by rich club regions in the context of functional network formation and global brain communication. This question might be addressed by examining large-scale computational models able to capture dynamics at distinct temporal scales as observed using simultaneous electroencephalography and fMRI recordings. The second question pertains to the relevance of rich club oscillations for the transitions between functional network configurations. While chapter 5 showed that rich club regions exhibit oscillations to varying degrees depending on task demands placed on the brain, it remains to be investigated to which extend transitions from rest to task performance or from one task to another are brought about by changes in the dynamics exhibited by rich club regions; that is by dynamically adjusting oscillatory behavior. The third, and final question proposed here concerns the behavioral relevance of rich club oscillations and functional network formation in general. Future research might address this question by moving the brain into states that differ from the state associated with a certain task to varying degrees and observe the temporal evolution of performance on that task as well as of local brain dynamics and of functional connectivity. Control of brain states could be achieved behaviorally in a task switching paradigm where tasks of varying degrees of similarity with the target task in terms of rich club dynamics or global functional connectivity need to be executed. Alternatively, transcranial magnetic stimulation (TMS), guided by large-scale computational models, might be used to move the brain towards or away from the brain state associated with the target task and observing the effect of these manipulations on task performance. Furthermore, using TMS to selectively target the rich club one might further test the idea that global brain states can be controlled by establishing control over the rich club.

A more general direction future research might take is to investigate interactions across spatial scales for a specific functional state of the brain (Lewis, Bosman, & Fries, 2015). From an empirical side, as a result of continuing developments with regard to ultra-high field fMRI, research might increasingly move towards whole-brain measurement with adequate spatial

resolution (e.g. De Martino et al., 2011; Gorgolewski et al., 2015) allowing for the simultaneous observation of global and region specific spatio-temporal activation profiles. From a theoretical side, computational neuroscientists might increasingly embed models locally implementing a specific mechanism, such as object recognition, into large-scale models of the cortex. Another potential avenue to elucidate interactions across scales is to investigate functional connectivity at the mesoscopic scale. To that end the single population neural mass models employed in currently used large-scale models would need to be replaced by units describing local dynamics to a larger degree of detail. One potential candidate would be the recently developed mean-field reduction of a layered cortical microcircuit (Bos, Diesmann, & Helias, 2015; Potjans & Diesmann, 2014; Schuecker, Helias, & Diesmann, 2016; Schuecker, Schmidt, van Albada, Diesmann, & Helias, 2015). Such a model could be used to simultaneously study the dynamics of global and local functional connectivity as well as their interaction.

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.
- Akaike, H. (1978). On the likelihood of a time series model. *The Statistician*, 217–235.
- Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*, 66(2), 237–242.
- Allen, J. F. (2001). Bioinformatics and discovery: induction beckons again. *BioEssays*, 23(1), 104–107. [http://doi.org/10.1002/1521-1878\(200101\)23:1<104::AID-BIES1013>3.0.CO;2-2](http://doi.org/10.1002/1521-1878(200101)23:1<104::AID-BIES1013>3.0.CO;2-2)
- Andersson, J. L. R., & Skare, S. (2002). A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI. *Neuroimage*, 16(1), 177–199.
- Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2), 870–888.
- Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupe, J.-M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *The Journal of Neuroscience*, 22(19), 8633–8646.
- Baars, B. J. (1993). *A cognitive theory of consciousness*. Cambridge University Press.
- Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. Oxford University Press on Demand.
- Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences*, 6(1), 47–52.
- Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *The Boundaries of Consciousness: Neurobiology and Neuropathology*, 150, 45–53.
- Balasubramanian, M., Polimeni, J., & Schwartz, E. L. (2002). The V1–V2–V3 complex: quasiconformal dipole maps in primate striate and extra-striate cortex. *Neural Networks*, 15(10), 1157–1163.
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509.
- Bassett, D. S., Wymbs, N. F., Rombach, M. P., Porter, M. A., Mucha, P. J., & Grafton, S. T. (2012). Core-Periphery Organisation of Human Brain Dynamics. *arXiv:1210.3555*.
- Bear, M. F., Connors, B., & Paradiso, M. (2006). *Neuroscience: Exploring the Brain* (3rd Revise). Lippincott Williams and Wilkins.
- Binda, P., Thomas, J. M., Boynton, G. M., & Fine, I. (2013). Minimizing biases in estimating the reorganization of human visual areas with BOLD retinotopic mapping. *Journal of Vision*, 13(7). <http://doi.org/10.1167/13.7.13>
- Blakeslee, B., Cope, D., & McCourt, M. E. (2015). The Oriented Difference of Gaussians (ODOG) model of brightness perception: Overview and executable Mathematica notebooks. *Behavior Research Methods*, 1–7.
- Blakeslee, B., & McCourt, M. E. (1999). A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. *Vision Research*, 39(26), 4361–4377.
- Blakeslee, B., & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Research*, 44(21), 2483–2503.
- Bonin, V., Mante, V., & Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus. *The Journal of Neuroscience*, 25(47), 10844–10856.
- Boone, W., & Piccinini, G. (2015). The cognitive neuroscience revolution. *Synthese*, 1–26. <http://doi.org/10.1007/s11229-015-0783-4>
- Bos, H., Diesmann, M., & Helias, M. (2015). Identifying anatomical origins of coexisting oscillations in the cortical microcircuit. *arXiv Preprint arXiv:1510.00642*.

References

- Braga, R. M., Sharp, D. J., Leeson, C., Wise, R. J., & Leech, R. (2013). Echoes of the Brain within Default Mode, Association, and Heteromodal Cortices. *The Journal of Neuroscience*, 33(35), 14031–14039.
- Breakspear, M., Jirsa, V., & Deco, G. (2010). Computational models of the brain: From structure to function. *NeuroImage*, 52(3), 727–730. <http://doi.org/10.1016/j.neuroimage.2010.05.061>
- Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth.
- Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5), 336–349.
- Cabral, J., Hugues, E., & Deco, G. (2011). Simulated functional networks in health and schizophrenia: a graph theoretical approach. *BMC Neuroscience*, 12(Suppl 1), P63.
- Cabral, J., Luchkoo, H., Woolrich, M., Joensson, M., Mohseni, H., Baker, A., ... Deco, G. (2014). Exploring mechanisms of spontaneous functional connectivity in MEG: how delayed network interactions lead to structured amplitude envelopes of band-pass filtered oscillations. *NeuroImage*, 90, 423–435.
- Campbell, F. W., & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, 197(3), 551–566.
- Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, 15(366), 92–96.
- Chatham, C. H., & Badre, D. (2014). How to test cognitive theory with fMRI. *arXiv:1404.2917 [Q-Bio]*.
- Chevreul, E. (1839). *De la loi du contraste simultané des couleurs et de l'assortiment des objets colorés*. Chez Pitois-Levrault.
- Chomsky, N. (1964). *Aspects of the Theory of Syntax*.
- Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. The MIT press.
- Clarke, K. A. (2007). The necessity of being comparative: Theory confirmation in quantitative political science. *Comparative Political Studies*.
- Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, 16(9), 1348–1355.
- Colizza, V., Flammini, A., Serrano, M. A., & Vespignani, A. (2006). Detecting rich-club ordering in complex networks. *Nature Physics*, 2(2), 110–115.
- Collin, G., Hulshoff Pol, H. E., Haijma, S. V., Cahn, W., Kahn, R. S., & van den Heuvel, M. P. (2011). Impaired Cerebellar Functional Connectivity in Schizophrenia Patients and Their Healthy Siblings. *Frontiers in Psychiatry*, 2. <http://doi.org/10.3389/fpsy.2011.00073>
- Collin, G., Sporns, O., Mandl, R. C., & van den Heuvel, M. P. (2013). Structural and Functional Aspects Relating to Cost and Benefit of Rich Club Organization in the Human Cerebral Cortex. *Cerebral Cortex*.
- Coltheart, M. (2006). What has Functional Neuroimaging told us about the Mind (so far)? <i>i>(Position Paper Presented to the European Cognitive Neuropsychology Workshop, Bressanone, 2005)</i>. *Cortex*, 42(3), 323–331.
- Condorcet, M. de. (1785). An essay on the application of probability theory to plurality decision making: An election between three candidates. *The Political Theory of Condorcet*, 69–80.
- Corchs, S., & Deco, G. (2002). Large-scale Neural Model for Visual Attention: Integration of Experimental Single-cell and fMRI Data. *Cerebral Cortex*, 12(4), 339–348. <http://doi.org/10.1093/cercor/12.4.339>
- Craft, E., Schütze, H., Niebur, E., & Heydt, R. von der. (2007). A Neural Model of Figure–Ground Organization. *Journal of Neurophysiology*, 97(6), 4310–4326. <http://doi.org/10.1152/jn.00203.2007>

- Dakin, S. C., & Bex, P. J. (2003). Natural image statistics mediate brightness “filling in.” *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1531), 2341–2348.
- Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33(1), 25–62.
- De Martino, F., Esposito, F., van de Moortele, P.-F., Harel, N., Formisano, E., Goebel, R., ... Yacoub, E. (2011). Whole brain high-resolution functional imaging at ultra high magnetic fields: an application to the analysis of resting state networks. *Neuroimage*, 57(3), 1031–1044. <http://doi.org/10.1016/j.neuroimage.2011.05.008>
- de Reus, M. A., & Heuvel, M. P. van den. (2013). Rich Club Organization and Intermodule Communication in the Cat Connectome. *The Journal of Neuroscience*, 33(32), 12929–12939. <http://doi.org/10.1523/JNEUROSCI.1448-13.2013>
- de Reus, M. A., & van den Heuvel, M. P. (2013). Rich Club Organization and Intermodule Communication in the Cat Connectome. *The Journal of Neuroscience*, 33(32), 12929–12939.
- de Reus, M. A., & van den Heuvel, M. P. (2014). Simulated rich club lesioning in brain networks: a scaffold for communication and integration? *Frontiers in Human Neuroscience*, 8. <http://doi.org/10.3389/fnhum.2014.00647>
- De Valois, R. L., Webster, M. A., De Valois, K. K., & Lingelbach, B. (1986). Temporal properties of brightness and color induction. *Vision Research*, 26(6), 887–897.
- Deco, G., & Jirsa, V. K. (2012). Ongoing Cortical Activity at Rest: Criticality, Multistability, and Ghost Attractors. *The Journal of Neuroscience*, 32(10), 3366–3375. <http://doi.org/10.1523/JNEUROSCI.2523-11.2012>
- Deco, G., Jirsa, V. K., & McIntosh, A. R. (2010). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, 12(1), 43–56.
- Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, 12(1), 43–56. <http://doi.org/10.1038/nrn2961>
- Deco, G., Senden, M., & Jirsa, V. (2012). How anatomy shapes dynamics: a semi-analytical study of the brain at rest by a simple spin model. *Frontiers in Computational Neuroscience*. <http://doi.org/10.3389/fn-com.2012.00068>
- Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79(1), 1–37.
- DeValois, R. L., DeValois, B. K. K., & others. (1988). *Spatial vision*. Oxford University Press, USA.
- DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., ... Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences*, 93(6), 2382–2386.
- DeYoe, E., Puckett, A., & Ma, Y. (2013). Population Attentional Field Modeling. *Journal of Vision*, 13(9), 232–232. <http://doi.org/10.1167/13.9.232>
- Dilks, D. D., Baker, C. I., Peli, E., & Kanwisher, N. (2009). Reorganization of visual processing in macular degeneration is not specific to the “preferred retinal locus.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(9), 2768–2773. <http://doi.org/10.1523/JNEUROSCI.5258-08.2009>
- Dumoulin, S. O., Hess, R. F., May, K. A., Harvey, B. M., Rokers, B., & Barendregt, M. (2014). Contour extracting networks in early extrastriate cortex. *Journal of Vision*, 14(5). <http://doi.org/10.1167/14.5.18>

References

- Dumoulin, S. O., Hoge, R. D., Baker Jr, C. L., Hess, R. F., Achtman, R. L., & Evans, A. C. (2003). Automatic volumetric segmentation of human visual retinotopic cortex. *Neuroimage*, *18*(3), 576–587.
- Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*(2), 647–660.
- Duyn, J. H. (2012). The future of ultra-high field MRI and fMRI for study of the human brain. *Neuroimage*, *62*(2), 1241–1248. <http://doi.org/10.1016/j.neuroimage.2011.10.065>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception & Psychophysics*, *16*(1), 143–149.
- Fisher, N. I., & Lee, A. J. (1983). A correlation coefficient for circular data. *Biometrika*, *70*(2), 327–332.
- Flourens, P. (1842). *Recherches expérimentales sur les propriétés et les fonctions du système nerveux dans les animaux vertébrés*. Ballière.
- Flowers, K. A., & Robertson, C. (1985). The effect of Parkinson's disease on the ability to maintain a mental set. *Journal of Neurology, Neurosurgery & Psychiatry*, *48*(6), 517–529.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201. <http://doi.org/10.1038/nn.2889>
- Freyer, F., Roberts, J. A., Becker, R., Robinson, P. A., Ritter, P., & Breakspear, M. (2011). Biophysical mechanisms of multistability in resting-state cortical rhythms. *The Journal of Neuroscience*, *31*(17), 6353–6361.
- Freyer, F., Roberts, J. A., Ritter, P., & Breakspear, M. (2012). A canonical model of multistability and scale-invariance in biological systems. *PLoS Comput Biol*, *8*(8), e1002634.
- Fries, P. (2001). Rapid feature selective neuronal synchronization through correlated latency shifting. *Nature Neuroscience*, *4*(2), 194.
- Fries, P. (2009). Neuronal Gamma-Band Synchronization as a Fundamental Process in Cortical Computation. *Annual Review of Neuroscience*, *32*(1), 209–224. <http://doi.org/10.1146/annurev.neuro.051508.135603>
- Frigg, R., & Hartmann, S. (2012). Models in Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2012).
- Friston, K. J., & Dolan, R. J. (2010). Computational and dynamic models in neuroimaging. *Neuroimage*, *52*(3), 752–765.
- Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *The Journal of Physiology*, *252*(3), 627–656.
- Ghosh, A., Rho, Y., McIntosh, A. R., Kötter, R., & Jirsa, V. K. (2008). Noise during rest enables the exploration of the brain's dynamic repertoire. *PLoS Comput Biol*, *4*(10), e1000196.
- Gibbs, J. W. (1905). *Elementare Grundlagen der statistischen Mechanik: entwickelt besonders im Hinblick auf eine rationelle Begründung der Thermodynamik*. JA Barth.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., ... others. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*, *80*, 105–124.
- Glerean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P., & Sams, M. (2012). Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity. *Brain Connectivity*, *2*(2), 91–101.
- Goebel, R., & De Weerd, P. (2009). Perceptual filling-in: from experimental data to neural network modeling. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences* (Vol. 6, pp. 435–456). Cambridge, MA: MIT Press.

- Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping, 27*(5), 392–401.
- Gollo, L. L., Zalesky, A., Hutchison, R. M., van den Heuvel, M., & Breakspear, M. (2015). Dwelling quietly in the rich club: brain network determinants of slow cortical fluctuations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 370*(1668), 20140165.
- Gorgolewski, K. J., Mendes, N., Wilfling, D., Wladimirow, E., Gauthier, C. J., Bonnen, T., ... others. (2015). A high resolution 7-Tesla resting-state fMRI test-retest dataset with cognitive and physiological measures. *Scientific Data, 2*.
- Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science, 7*(2), 99–108.
- Grigg, O., & Grady, C. L. (2010). Task-related effects on the temporal and spatial dynamics of resting-state functional connectivity in the default network. *PLoS One, 5*(10), e13311–e13311.
- Grossberg, S. (2003). Filling-In the Forms: Surface and Boundary Interactions. *Filling-In: From Perceptual Completion to Cortical Reorganization: From Perceptual Completion to Cortical Reorganization*, 13.
- Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Alfred, B. Y., Kahn, A. E., ... others. (2015). Controllability of structural brain networks. *Nature Communications, 6*.
- Haak, K. V., Langers, D. R. M., Renken, R., van Dijk, P., Borgstein, J., & Cornelissen, F. W. (2014). Abnormal visual field maps in human cortex: A mini-review and a case report. *Cortex, 0*(0). <http://doi.org/10.1016/j.cortex.2012.12.005>
- Harriger, L., van den Heuvel, M. P., & Sporns, O. (2012). Rich club organization of macaque cerebral cortex and its role in network communication. *PLoS One, 7*(9), e46497.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*(1), 97–109.
- Havlicek, M., Roebroeck, A., Friston, K., Gardumi, A., Ivanov, D., & Uludag, K. (2015). Physiologically informed dynamic causal modeling of fMRI data. *NeuroImage, 122*, 355–372. <http://doi.org/10.1016/j.neuroimage.2015.07.078>
- Hedges, L. V., Olkin, I., & Statistiker, M. (1985). *Statistical methods for meta-analysis*. Academic Press New York.
- Henriksson, L., Nurminen, L., Hyvärinen, A., & Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. *Journal of Vision, 8*(10), 5.
- Hess, C., & Pretori, H. (1894). Messende Untersuchungen über die Gesetzmässigkeit des simultanen Helligkeits-contrastes. *Graefe's Archive for Clinical and Experimental Ophthalmology, 40*(4), 1–24.
- Hiltunen, T., Kantola, J., Elseoud, A. A., Lepola, P., Suominen, K., Starck, T., ... Palva, J. M. (2014). Infra-Slow EEG Fluctuations Are Correlated with Resting-State Network Dynamics in fMRI. *The Journal of Neuroscience, 34*(2), 356–362. <http://doi.org/10.1523/JNEUROSCI.0276-13.2014>
- Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., & Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience, 15*(6), 884–890.
- Hoffman, R. E., Gueorguieva, R., Hawkins, K. A., Varanko, M., Boutros, N. N., Wu, Y., ... Krystal, J. H. (2005). Temporoparietal Transcranial Magnetic Stimulation for Auditory Hallucinations: Safety, Efficacy and Moderators in a Fifty Patient Sample. *Biological Psychiatry, 58*(2), 97–104. <http://doi.org/10.1016/j.biopsych.2005.03.041>
- Holmes, G. (1918). Disturbances of vision by cerebral lesions. *The British Journal of Ophthalmology, 2*(7), 353.

References

- Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., & Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, *106*(6), 2035.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79*(8), 2554–2558.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591.
- Hurlbert, A. C., & Poggio, T. A. (1988). Synthesizing a color algorithm from examples. *Science*, *239*(4839), 482–485.
- Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., ... others. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, *80*, 360–378.
- Jeffrey, P. (2003). Smoothing the Waters Observations on the Process of Cross-Disciplinary Research Collaboration. *Social Studies of Science*, *33*(4), 539–562. <http://doi.org/10.1177/0306312703334003>
- Jehee, J. F. M., Lamme, V. A. F., & Roelfsema, P. R. (2007). Boundary assignment in a recurrent network architecture. *Vision Research*, *47*(9), 1153–1165.
- Johnston, S. J., Boehm, S. G., Healy, D., Goebel, R., & Linden, D. E. J. (2010). Neurofeedback: A promising tool for the self-regulation of emotion networks. *NeuroImage*, *49*(1), 1066–1072. <http://doi.org/10.1016/j.neuroimage.2009.07.056>
- Jones, H. E., Grieve, K. L., Wang, W., & Sillito, A. M. (2001). Surround Suppression in Primate V1. *Journal of Neurophysiology*, *86*(4), 2011–2028.
- Kant, I. (1787). „Kritik der reinen Vernunft (2. Auflage)“. *Kant's Gesammelte Schriften*.
- Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport*, *10*(1), 183–187.
- Kay, K. N., Winawer, J., Mezer, A., & Wandell, B. A. (2013). Compressive spatial summation in human visual cortex. *Journal of Neurophysiology*, *110*(2), 481–494.
- Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLoS Comput Biol*, *10*(11), e1003915. <http://doi.org/10.1371/journal.pcbi.1003915>
- Kingdom, F. A. A. (2011). Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, *51*(7), 652–673.
- Kinoshita, M., & Komatsu, H. (2001). Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *Journal of Neurophysiology*, *86*(5), 2559–2570.
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352.
- Kriegeskorte, N., & Goebel, R. (2001). An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. *NeuroImage*, *14*(2), 329–346.
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.
- Krohs, U. (2012). Convenience experimentation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 52–57. <http://doi.org/10.1016/j.shpsc.2011.10.005>
- Lakatos, I. (1976). *Falsification and the methodology of scientific research programmes*. Springer.
- Landau, L. D., & Lifšic, E. M. (2007). *Course of theoretical physics. 5, 5.*, Oxford [u.a.]: Pergamon Press.
- Lashley, K. S. (1929). Brain mechanisms and intelligence: A quantitative study of injuries to the brain.

- Le Van Quyen, M. (2011). The brainweb of cross-scale interactions. *New Ideas in Psychology, 29*(2), 57–63. <http://doi.org/10.1016/j.newideapsych.2010.11.001>
- Lee, S., Papanikolaou, A., Logothetis, N. K., Smirnakis, S. M., & Keliris, G. A. (2013). A new method for estimating population receptive field topography in visual cortex. *NeuroImage*.
- Leech, R., Braga, R., & Sharp, D. J. (2012). Echoes of the brain within the posterior cingulate cortex. *The Journal of Neuroscience, 32*(1), 215–222.
- Leonelli, S. (2012). Introduction: Making sense of data-driven research in the biological and biomedical sciences. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43*(1), 1–3. <http://doi.org/10.1016/j.shpsc.2011.10.001>
- Levins, R. (2006). Strategies of abstraction. *Biology and Philosophy, 21*(5), 741–755.
- Lewis, C. M., Bosman, C. A., & Fries, P. (2015). Recording of brain activity across spatial scales. *Current Opinion in Neurobiology, 32*, 68–77. <http://doi.org/10.1016/j.conb.2014.12.007>
- Li, W., & Li, C.-Y. (1993). Integration field beyond the classical visual receptive field. *Chinese Journal of Neuroscience, 1*(2), 1–6.
- Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature, 473*(7346), 167–173. <http://doi.org/10.1038/nature10011>
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150–157.
- Lorente de Nó, R. (1949). Cerebral cortex: architecture, intracortical connections, motor projections. *Physiology of the Nervous System, 288–330*.
- MacEvoy, S. P., Kim, W., & Paradiso, M. A. (1998). Integration of surface information in primary visual cortex. *Nature Neuroscience, 1*(7), 616–620.
- Maier, A., Wilke, M., Aura, C., Zhu, C., Frank, Q. Y., & Leopold, D. A. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nature Neuroscience, 11*(10), 1193–1200.
- Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., & Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T 1-mapping at high field. *NeuroImage, 49*(2), 1271–1281.
- Marr, D. (1977). Artificial intelligence—a personal view. *Artificial Intelligence, 9*(1), 37–48.
- Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information, henry holt and co. *Inc., New York, NY, 2*.
- Marr, D., & Poggio, T. (1976). From understanding computation to understanding neural circuitry.
- Maslov, S., & Sneppen, K. (2002). Specificity and Stability in Topology of Protein Networks. *Science, 296*(5569), 910–913. <http://doi.org/10.1126/science.1065103>
- Mather, M., Cacioppo, J. T., & Kanwisher, N. (2013). How fMRI can inform cognitive theories. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 8*(1), 108–113. <http://doi.org/10.1177/1745691612469037>
- McAuley, J. J., da Fontoura Costa, L., & Caetano, T. S. (2007). Rich-club phenomenon across complex network hierarchies. *Applied Physics Letters, 91*(8), 084103–084103.
- Meng, M., Remus, D. A., & Tong, F. (2005). Filling-in of visual phantoms in the human brain. *Nature Neuroscience, 8*(9), 1248–1254.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics, 21*, 1087.

References

- Michalareas, G., Vezioli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. *Neuron*.
- Monto, S., Palva, S., Voipio, J., & Palva, J. M. (2008). Very Slow EEG Fluctuations Predict the Dynamics of Stimulus Detection and Oscillation Amplitudes in Humans. *The Journal of Neuroscience*, 28(33), 8268–8272. <http://doi.org/10.1523/JNEUROSCI.1910-08.2008>
- Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain*, 120(4), 701–722.
- Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978a). Receptive field organization of complex cells in the cat's striate cortex. *The Journal of Physiology*, 283, 79.
- Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of Physiology*, 283, 53.
- Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., ... Yacoub, E. (2015). Contextual feedback to superficial layers of V1. *Current Biology*, 25(20), 2690–2695.
- Nageswaran, J. M., Dutt, N., Krichmar, J. L., Nicolau, A., & Veidenbaum, A. V. (2009). A configurable simulation environment for the efficient simulation of large-scale spiking neural networks on graphics processors. *Neural Networks*, 22(5-6), 791–800.
- O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, 401(6753), 584–587. <http://doi.org/10.1038/44134>
- Page, M. (2006). What can't functional neuroimaging tell the cognitive psychologist? *Cortex*, 42(3), 428–443.
- Patel, A. X., & Bullmore, E. T. (2015). A wavelet-based estimator of the degrees of freedom in denoised fMRI time series for probabilistic testing of functional connectivity and brain graphs. *NeuroImage*.
- Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, 45(1), S199–S209.
- Peters, J. C., Jans, B., van de Ven, V., De Weerd, P., & Goebel, R. (2010). Dynamic brightness induction in V1: Analyzing simulated and empirically acquired fMRI data in a "common brain space" framework. *Neuroimage*, 52(3), 973–984.
- Peters, J. C., Reithler, J., & Goebel, R. (2012). Modeling invariant object processing based on tight integration of simulated and empirical data in a Common Brain Space. *Frontiers in Computational Neuroscience*, 6.
- Platt, J. R. (1964). Strong inference. *Science*, 146(3642), 347–353.
- Polimeni, J. R., Hinds, O. P., Balasubramanian, M., van der Kouwe, A., Wald, L. L., Dale, A. M., ... others. (2005). The human V1–V2–V3 visuotopic map complex measured via fMRI at 3 and 7 Tesla.
- Poser, B. A., Koopmans, P. J., Witzel, T., Wald, L. L., & Barth, M. (2010). Three dimensional echo-planar imaging at 7 Tesla. *Neuroimage*, 51(1), 261–266.
- Potjans, T. C., & Diesmann, M. (2014). The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. *Cerebral Cortex*, 24(3), 785–806.
- Rescorla, M. (2015). The Computational Theory of Mind. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015).
- Robinson, A. E., Hammon, P. S., & de Sa, V. R. (2007). Explaining brightness illusions using spatial filtering and local response normalization. *Vision Research*, 47(12), 1631–1644.
- Roelfsema, P. R., Lamme, V. A. F., Spekreijse, H., & Bosch, H. (2002). Figure—ground segregation in a recurrent network architecture. *Journal of Cognitive Neuroscience*, 14(4), 525–537.
- Rolls, E. T., Grabenhorst, F., & Deco, G. (2010). Decision-making, errors, and confidence in the brain. *Journal of Neurophysiology*, 104(5), 2359–2374.

- Rossi, A. F., Rittenhouse, C. D., & Paradiso, M. A. (1996). The representation of brightness in primary visual cortex. *Science*, *273*(5278), 1104–1107.
- Rudd, M. E. (2001). Lightness computation by a neural filling-in mechanism. In *Photonics West 2001-Electronic Imaging* (pp. 400–413). International Society for Optics and Photonics.
- Rudd, M. E., & Arrington, K. F. (2001). Darkness filling-in: a neural model of darkness induction. *Vision Research*, *41*(27), 3649–3662.
- Rudd, M. E., & Popa, D. (2007). Stevens's brightness law, contrast gain control, and edge integration in achromatic color perception: a unified model. *JOSA A*, *24*(9), 2766–2782.
- Sack, A. T., Kadosh, R. C., Schuhmann, T., Moerel, M., Walsh, V., & Goebel, R. (2009). Optimizing functional accuracy of TMS in cognitive studies: a comparison of methods. *Journal of Cognitive Neuroscience*, *21*(2), 207–221.
- Sasaki, Y., Rajimehr, R., Kim, B. W., Ekstrom, L. B., Vanduffel, W., & Tootell, R. B. H. (2006). The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron*, *51*(5), 661–670.
- Schmidt, R., LaFleur, K. J. R., de Reus, M. A., van den Berg, L. H., & van den Heuvel, M. P. (2015). Kuramoto model simulation of neural hubs and dynamic synchrony in the human cerebral connectome. *BMC Neuroscience*, *16*(1), 1.
- Schuecker, J., Helias, M., & Diesmann, M. (2016). Mean-field reduction of a layered macroscopic network model. Schuecker, J., Schmidt, M., van Albada, S., Diesmann, M., & Helias, M. (2015). Fundamental activity constraints lead to specific interpretations of the connectome. *arXiv Preprint arXiv:1509.03162*.
- Schwartz, E. L. (1980). Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision Research*, *20*(8), 645–669.
- Schwarz, G., & others. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.
- Senden, M., Deco, G., de Reus, M. A., Goebel, R., & van den Heuvel, M. P. (2014). Rich club organization supports a diverse set of functional network configurations. *NeuroImage*, *96*, 174–182. <http://doi.org/10.1016/j.neuroimage.2014.03.066>
- Senden, M., Goebel, R., & Deco, G. (2012). Structural connectivity allows for multi-threading during rest: The structure of the cortex leads to efficient alternation between resting state exploratory behavior and default mode processing. *NeuroImage*, *60*(4), 2274–2284.
- Senden, M., Reithler, J., Gijsen, S., & Goebel, R. (2014). Evaluating Population Receptive Field Estimation Frameworks in Terms of Robustness and Reproducibility. *PLoS ONE*, *9*(12), e114054. <http://doi.org/10.1371/journal.pone.0114054>
- Sereno, M., Dale, A., Reppas, J., Kwong, K., Belliveau, J., Brady, T., ... Tootell, R. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889.
- Shanahan, M. (2012). The brain's connective core and its role in animal cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1603), 2704–2714. <http://doi.org/10.1098/rstb.2012.0128>
- Shapiro, A., Knight, E., & Lu, Z.-L. (2010). Spatial scale models of lightness illusions: contrast, anchoring, and tunable filters. *Journal of Vision*, *8*(6), 288. <http://doi.org/10.1167/8.6.288>
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects.
- Shmuel, A., Yacoub, E., Chaimow, D., Logothetis, N. K., & Ugurbil, K. (2007). Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. *NeuroImage*, *35*(2), 539–552. <http://doi.org/10.1016/j.neuroimage.2006.12.030>

References

- Silberberg, G., Grillner, S., LeBeau, F. E. N., Maex, R., & Markram, H. (2005). Synaptic pathways in neural microcircuits. *Trends in Neurosciences*, *28*(10), 541–551.
- Sperry, R. W. (1963). Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proceedings of the National Academy of Sciences*, *50*(4), 703–710.
- Steiger, J. H. (2003). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*(2), 164–182.
- Szentágothai, J. (1983). The modular architectonic principle of neural centers. In *Reviews of Physiology, Biochemistry and Pharmacology, Volume 98* (pp. 11–61). Springer.
- Tagamets, M. A., & Horwitz, B. (1998). Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. *Cerebral Cortex*, *8*(4), 310–320. <http://doi.org/10.1093/cercor/8.4.310>
- Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., LeBihan, D., & Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, *33*(4), 1104–1116.
- Tideman, T. N. (1987). Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, *4*(3), 185–206.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5*(1), 42.
- Towson, E. K., Vértés, P. E., Ahnert, S. E., Schafer, W. R., & Bullmore, E. T. (2013). The rich club of the *C. elegans* neuronal connectome. *The Journal of Neuroscience*, *33*(15), 6380–6387.
- van de Ven, V., Jans, B., Goebel, R., & De Weerd, P. (2011). Early Human Visual Cortex Encodes Surface Brightness Induced by Dynamic Context. *Journal of Cognitive Neuroscience*, *24*(2), 367–377. http://doi.org/10.1162/jocn_a_00126
- van den Heuvel, M. P., Kahn, R. S., Goñi, J., & Sporns, O. (2012). High-cost, high-capacity backbone for global brain communication. *Proceedings of the National Academy of Sciences*, *109*(28), 11372–11377.
- van den Heuvel, M. P., & Sporns, O. (2011). Rich-club organization of the human connectome. *The Journal of Neuroscience*, *31*(44), 15775–15786.
- van den Heuvel, M. P., & Sporns, O. (2013a). An Anatomical Substrate for Integration among Functional Networks in Human Cortex. *The Journal of Neuroscience*, *33*(36), 14489–14500.
- van den Heuvel, M. P., & Sporns, O. (2013b). Network hubs in the human brain. *Trends in Cognitive Sciences*, *17*(12), 683–696. <http://doi.org/10.1016/j.tics.2013.09.012>
- van den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Pol, H. E. H. (2009). Efficiency of functional brain networks and intellectual performance. *The Journal of Neuroscience*, *29*(23), 7619–7624.
- Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., ... others. (2012). The Human Connectome Project: a data acquisition perspective. *Neuroimage*, *62*(4), 2222–2231.
- Van Horn, J. D., & Toga, A. W. (2014). Human neuroimaging as a “Big Data” science. *Brain Imaging and Behavior*, *8*(2), 323–331. <http://doi.org/10.1007/s11682-013-9255-y>
- Vanhatalo, S., Palva, J. M., Holmes, M. D., Miller, J. W., Voipio, J., & Kaila, K. (2004). Infralow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(14), 5053–5057.
- Váša, F., Shanahan, M., Hellyer, P. J., Scott, G., Cabral, J., & Leech, R. (2015). Effects of lesions on synchrony and metastability in cortical networks. *NeuroImage*, *118*, 456–467. <http://doi.org/10.1016/j.neuroimage.2015.05.042>

- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196.
- Walker, G. A., Ohzawa, I., & Freeman, R. D. (1999). Asymmetric Suppression Outside the Classical Receptive Field of the Visual Cortex. *The Journal of Neuroscience*, *19*(23), 10536–10553.
- Wandell, B. A. (1999). Computational Neuroimaging of Human Visual Cortex. *Annual Review of Neuroscience*, *22*(1), 145–173. <http://doi.org/10.1146/annurev.neuro.22.1.145>
- Wandell, B. A., & Winawer, J. (2015). Computational neuroimaging and population receptive fields. *Trends in Cognitive Sciences*.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, *393*(6684), 440–442.
- Whewell, W. (1858). *History of scientific ideas* (Vol. 1). JW Parker and son.
- Wildie, M., & Shanahan, M. (2012). Metastability and chimera states in modular delay and pulse-coupled oscillator networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *22*(4), 43131.
- Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience*, *26*(4), 1314–1328.
- Woodman, M. M., & Jirsa, V. K. (2013). Accelerating The Virtual Brain with code generation and GPU computing. *BMC Neuroscience*, *14*(Suppl 1), P198. <http://doi.org/10.1186/1471-2202-14-S1-P198>
- Yacoub, E., Harel, N., & Uğurbil, K. (2008). High-field fMRI unveils orientation columns in humans. *Proceedings of the National Academy of Sciences*, *105*(30), 10607–10612.
- Yacoub, E., Shmuel, A., Logothetis, N., & Uğurbil, K. (2007). Robust detection of ocular dominance columns in humans using Hahn Spin Echo BOLD functional MRI at 7 Tesla. *Neuroimage*, *37*(4), 1161–1177.
- Yacubian, J., Sommer, T., Schroeder, K., Gläscher, J., Braus, D. F., & Büchel, C. (2007). Subregions of the ventral striatum show preferential coding of reward magnitude and probability. *Neuroimage*, *38*(3), 557–563.
- Yeh, F.-C., Wedeen, V. J., & Tseng, W.-Y. I. (2010). Generalized-sampling imaging. *Medical Imaging, IEEE Transactions on*, *29*(9), 1626–1635.
- Yoshioka, T., Blasdel, G. G., Levitt, J. B., & Lund, J. S. (1996). Relation between Patterns of Intrinsic Lateral Connectivity, Ocular Dominance, and Cytochrome Oxidase-Reactive Regions in Macaque Monkey Striate Cortex. *Cerebral Cortex*, *6*(2), 297–310. <http://doi.org/10.1093/cercor/6.2.297>
- Yu, H.-H., Verma, R., Yang, Y., Tibballs, H. A., Lui, L. L., Reser, D. H., & Rosa, M. G. P. (2010). Spatial and temporal frequency tuning in striate cortex: functional uniformity and specializations related to receptive field eccentricity. *European Journal of Neuroscience*, *31*(6), 1043–1062.
- Yu, Q., Sui, J., Liu, J., Plis, S. M., Kiehl, K. A., Pearlson, G., & Calhoun, V. D. (2012). Disrupted correlation between low frequency power and connectivity strength of resting state brain networks in schizophrenia. *Schizophrenia Research*.
- Zamora-López, G., Zhou, C., & Kurths, J. (2009). Graph analysis of cortical networks reveals complex anatomical communication substrate. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *19*(1), 15117.
- Zamora-López, G., Zhou, C., & Kurths, J. (2011). Exploring brain function from anatomical connectivity. *Frontiers in Neuroscience*, *5*.
- Zhou, S., & Mondragon, R. J. (2004). The rich-club phenomenon in the Internet topology. *IEEE Communications Letters*, *8*(3), 180 – 182. <http://doi.org/10.1109/LCOMM.2004.823426>
- Zimmermann, J., Goebel, R., De Martino, F., van de Moortele, P.-F., Feinberg, D., Adriany, G., ... Yacoub, E. (2011). Mapping the organization of axis of motion selective features in human area MT using high-field fMRI. *PLoS One*, *6*(12), e28716.

References

- Zuiderbaan, W., Harvey, B. M., & Dumoulin, S. O. (2012). Modeling center-surround configurations in population receptive fields using fMRI. *Journal of Vision*, 12(3), 10.

Appendices

Appendix

APPENDIX A1: STIMULATION VIDEOS

Illustrative visualizations of stimulus presentations for the first subjects can be viewed online.

PW: pRF_2014_supplementary

Wedge presented in orderly sequence

Wedge presented in random sequence

Ring presented in orderly sequence

Ring presented in random sequence

Bar presented in orderly sequence

Bar presented in random sequence

APPENDIX A2: EVALUATION OF PRF ESTIMATION PROCEDURES FROM SIMULATIONS OF V1 TYPE RECEPTIVE FIELDS

Here we present the ranking for simulated data with underlying V1 type receptive fields with the aim to investigate which stimulus type either interspersed by mean luminance (ML) periods or not (\neg ML) most faithfully recovers the known population receptive field parameters. To this end we obtained the similarity between original and recovered pRF parameters. Each stimulus was presented both with and without mean luminance periods in two separate runs. See Table A2.1.

Applying the Tideman method (Tideman, 1987) we obtained the winners of each pairwise comparison and their orderings. Subsequently, we obtained the ranking of the combinations by locking the results in a directed graph. The resulting graph had two sources since both the bar and WR_{post} stimuli presented without periods of mean luminance had outgoing but no incoming connections. With an indegree of 2 WR_{pre} presented interspersed with mean luminance periods clearly constitutes the Condorcet loser. Finally, all remaining combinations of stimulus and presence or absence of mean luminance periods had presented both an in- and outdegree of zero and are thus interchangeable.

Based on the rankings of the combinations of stimulus type and stimulus presentation we derived matrices of voting results for type and continuance, respectively. Table A2.2a shows the voting results with regard to stimulus. Since both bar and WR_{post} are ranked above WR_{pre} while there is no distinction between the two, there is no clear Condorcet winner for these simulations. With regard to continuance, table A2.2b shows the voting results for the comparison of presentations with and without mean luminance periods. Since for the majority of rankings the absence of mean luminance periods was preferable to their presence, the exclusion of mean luminance periods constituted the Condorcet winner for the simulations of V1 type receptive fields.

Table A2.1: Ranking for pRF estimate Similarity

A. Voting results for within-session explanatory power

		WR _{pre}		WR _{post}		Bar	
		orderly	random	orderly	random	orderly	random
WR _{pre}	orderly		0.5	0.4	0.3	0.2	0.2
	random	0.6		1.1	0.8	0.4	0.4
WR _{post}	orderly	0.0	0.5		0.0	0.2	0.2
	random	0.3	0.7	0.5		0.2	0.2
Bar	orderly	1.1	1.2	1.6	1.2		0.0
	random	1.6	1.6	2.0	1.5	0.3	

C. Directed graph for within-session explanatory power

		WR _{pre}		WR _{post}		Bar	
		orderly	random	orderly	random	orderly	random
WR _{pre}	orderly	0	0	1	0	0	0
	random	1	0	1	1	0	0
WR _{post}	orderly	0	0	0	0	0	0
	random	0	0	1	0	0	0
Bar	orderly	1	1	1	1	0	0
	random	1	1	1	1	1	0
Indegree		3	2	5	3	1	0

B. Pairwise comparisons for within-session explanatory power

Pair	Order
WR _{pre} , orderly ($g = 0.5$) vs. WR _{pre} , random ($g = 0.6$)	11
WR_{pre}, orderly ($g = 0.4$) vs. WR _{post} , orderly ($g = 0$)	13
WR _{pre} , orderly ($g = 0.3$) vs. WR _{post} , random ($g = 0.3$)	15
WR _{pre} , orderly ($g = 0.2$) vs. Bar, orderly ($g = 1.1$)	8
WR _{pre} , orderly ($g = 0.2$) vs. Bar, random ($g = 1.6$)	2
WR_{pre}, random ($g = 1.1$) vs. WR _{post} , orderly ($g = 0.5$)	9
WR_{pre}, random ($g = 0.8$) vs. WR _{post} , random ($g = 0.7$)	10
WR _{pre} , random ($g = 0.4$) vs. Bar, orderly ($g = 1.2$)	7
WR _{pre} , random ($g = 0.4$) vs. Bar, random ($g = 1.6$)	4
WR _{post} , orderly ($g = 0$) vs. WR_{post}, random ($g = 0.5$)	12
WR _{post} , orderly ($g = 0.2$) vs. Bar, orderly ($g = 1.6$)	3
WR _{post} , orderly ($g = 0.2$) vs. Bar, random ($g = 2$)	1
WR _{post} , random ($g = 0.2$) vs. Bar, orderly ($g = 1.2$)	6
WR _{post} , random ($g = 0.2$) vs. Bar, random ($g = 1.5$)	5
Bar, orderly ($g = 0$) vs. Bar, random ($g = 0.3$)	14

A) Tally: Each cell represents the sum of hedges' g values for the pairwise comparison of row over column obtained for each combination of subject and session. Since three subjects were tested in two sessions (each subject casted two votes), there were six votes in total. Only significant results were summed, i.e. those hedges' g values whose confidence intervals did not include zero. If zero lay within the confidence interval the vote was counted as indifferent between the two options. **B) Sort:** The first column shows the comparison of each pair including the hedges' g values of each winning over the other. The winner of each pair is highlighted in bold. The second column shows the order in which pairs were locked based on majorities starting with the largest. **C) Lock:** The graph depicts binary edges leading from row to column as well as the sum total of incoming edges for each vertex (stimulus-presentation combination).



Table A2.2: Voting results for pRF estimate Similarity split for stimulus and mode of presentation

A. Stimulus Type

	WR _{pre}	WR _{post}	Bar
WR _{pre}		0	0
WR _{post}	1		0
Bar	1	0	

B. Continuance

	ML	¬ML
ML		2
¬ML	0	

A) Stimulus type: Each cell represents the sum of rankings where the row stimulus was ranked above the column stimulus. **B) Mode of presentation:** Each cell represents the sum of rankings where the mode of presentation given by the row was ranked above the mode of presentation given by the column.

APPENDIX A3: EVALUATION OF PRF ESTIMATION PROCEDURES FROM SIMULATIONS OF V4 TYPE RECEPTIVE FIELDS

Here we present the ranking for simulated data with underlying V4 type receptive fields with the aim to investigate which stimulus type either interspersed by mean luminance (ML) periods or not (\neg ML) most faithfully recovers the known population receptive field parameters. To this end we obtained the similarity between original and recovered pRF parameters. Each stimulus was presented both with and without mean luminance periods in two separate runs. Table A3.1 shows the results. Applying the Tideman method (Tideman, 1987) we obtained the winners of each pairwise comparison as well as their orderings. Subsequently, we obtained the ranking of the combinations by locking the results in a directed graph. The adjacency matrix of this graph as well as each pair's indegree is represented in table A3.1c. Edges are directed from row to column. The source of the resulting graph was the bar presented interspersed by periods of mean luminance. With an indegree of 1, WR_{pre} presented interspersed with mean luminance periods as well as the bar presented without mean luminance periods were ranked second. WR_{pre} without mean luminance periods was ranked third with an indegree of 2. WR_{post} with mean luminance periods was ranked fourth with an indegree of 3. Finally, WR_{post} without mean luminance periods was ranked fifth with an indegree of 4.

Based on the rankings of the combinations of stimulus type and continuance we derived matrices of voting results for each of them separately. Table A3.2a shows the voting results with regard to stimulus. Since the bar was ranked consistently above the other stimuli, it constituted the Condorcet winner for these simulations. With regard to continuance, table A3.2b shows the voting results for the comparison of presentations with and without interspersed mean luminance periods. Since for the majority of rankings the presence of mean luminance periods was preferable to their absence, the inclusion of mean luminance periods constituted the Condorcet winner for the simulations.

Table A3.1: Ranking for pRF estimate Similarity

A. Voting results for pRF estimate Similarity

		WR _{pre}		WR _{post}		Bar	
		ML	¬ML	ML	¬ML	ML	¬ML
WR _{pre}	ML		0.0	0.4	0.5	0.0	0.0
	¬ML	0.0		0.0	0.3	0.0	0.0
WR _{post}	ML	0.0	0.0		0.0	0.0	0.0
	¬ML	0.0	0.0	0.0		0.0	0.0
Bar	ML	0.6	0.9	1.1	1.2		0.5
	¬ML	0.0	0.3	0.4	0.6	0.0	

B. Pairwise comparisons for pRF estimate Similarity

Pair	Order
WR _{pre} , ML (g = 0) vs. WR _{pre} , ¬ML (g = 0)	12
WR_{pre}, ML (g = 0.4) vs. WR _{post} , ML (g = 0)	8
WR_{pre}, ML (g = 0.5) vs. WR _{post} , ¬ML (g = 0)	6
WR _{pre} , ML (g = 0) vs. Bar , ML (g = 0.6)	4
WR _{pre} , ML (g = 0) vs. Bar , ¬ML (g = 0)	13
WR _{pre} , ¬ML (g = 0) vs. WR _{post} , ML (g = 0)	14
WR_{pre}, ¬ML (g = 0.3) vs. WR _{post} , ¬ML (g = 0)	10
WR _{pre} , ¬ML (g = 0) vs. Bar , ML (g = 0.9)	3
WR _{pre} , ¬ML (g = 0) vs. Bar , ¬ML (g = 0.3)	11
WR _{post} , ML (g = 0) vs. WR _{post} , ¬ML (g = 0)	15
WR _{post} , ML (g = 0) vs. Bar , ML (g = 1.1)	2
WR _{post} , ML (g = 0) vs. Bar , ¬ML (g = 0.4)	9
WR _{post} , ¬ML (g = 0) vs. Bar , ML (g = 1.2)	1
WR _{post} , ¬ML (g = 0) vs. Bar , ¬ML (g = 0.6)	5
Bar , ML (g = 0.5) vs. Bar , ¬ML (g = 0)	7

C. Directed Graph for pRF estimate Similarity

		WR _{pre}		WR _{post}		Bar	
		ML	¬ML	ML	¬ML	ML	¬ML
WR _{pre}	ML	0	0	1	1	0	0
	¬ML	0	0	0	1	0	0
WR _{post}	ML	0	0	0	0	0	0
	¬ML	0	0	0	0	0	0
Bar	ML	1	1	1	1	0	1
	¬ML	0	1	1	1	0	0
Indegree		1	2	3	4	0	1

A) Tally: Each cell represents the sum of hedges' g values for the pairwise comparison of row over column obtained for each combination of subject and session. Since three subjects were tested in two sessions (each subject casted two votes), there were six votes in total. Only significant results were summed, i.e. those hedges' g values whose confidence intervals did not include zero. If zero lay within the confidence interval the vote was counted as indifferent between the two options. **B) Sort:** The first column shows the comparison of each pair including the hedges' g values of each winning over the other. The winner of each pair is highlighted in bold. The second column shows the order in which pairs were locked based on majorities starting with the largest. **C) Lock:** The graph depicts binary edges leading from row to column as well as the sum total of incoming edges for each vertex (stimulus-presentation combination).

Table A3.2: Voting results for pRF estimate Similarity split for stimulus and mode of presentation

A. Stimulus Type

	WR _{pre}	WR _{post}	Bar
WR _{pre}		4	0
WR _{post}	0		0
Bar	3	4	

B. Continuance

	ML	¬ML
ML		5
¬ML	1	

A) Stimulus type: Each cell represents the sum of rankings where the row stimulus was ranked above the column stimulus. **B) Mode of presentation:** Each cell represents the sum of rankings where the mode of presentation given by the row was ranked above the mode of presentation given by the column.

APPENDIX A4: METROPOLIS HASTINGS ALGORITHM

In what follows the Metropolis Hastings Algorithm used to estimate the entropy in a spin glass neural network is described. First, an initial global configuration was chosen by randomly selecting the number of spins that would initially be in an 'up' state from a normal distribution whose mean was 50% of spins in an 'up' and 50% in a 'down' state and whose standard deviation was 12.5%. Subsequently, spins were randomly selected and switched to an 'up' state until the number of spins supposed to be 'up' was reached. From this initial condition the algorithm proceeded by randomly selecting one spin and flipping it. If the energy of a thusly selected trial configuration was lower as compared to the original configuration, the trial configuration was accepted. If the energy of the trial configuration was higher as compared to the original configuration, the trial configuration was accepted only with a certain probability:

$$P_{accept} = e^{-\varepsilon\delta H}$$

A4.1

where δH is the difference between the energy of the trial configuration and the energy of the original configuration: $\delta H = H^t - H^o$.

These described steps were repeated for one million iterations for matrices comprising $N = 30$ nodes and ten million iterations for matrices comprising $N = 68$ nodes. In order to avoid the algorithm from getting stuck in local minima, occasionally (every 500 iterations), all spins were concurrently flipped such that every spin in an 'up' state switched to a 'down' state and vice versa. After the algorithm was finished, all unique global configurations that had been visited were retained. The entropy was calculated as described in the method section but only over the set of unique global configurations as obtained from running the algorithm.

APPENDIX A5: SUPPLEMENTARY FIGURES CHAPTER 4

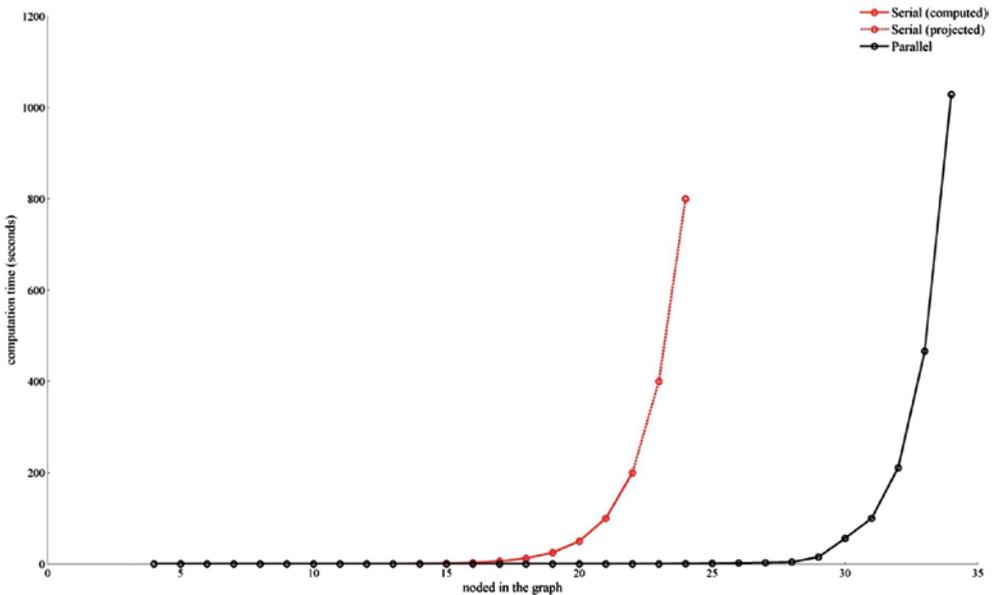


Figure A5.1: Computation Time. The figure shows the time needed for solving the spin glass model for a single network as a function of its order (i.e. its number of nodes). The solid red line indicates the actual computation time if computations of states are run in serial. Computation time increases twofold with the addition of a node. The dashed red line indicates a projection of the computation time for networks of larger size if computations are run in serial. The black line indicates the computation time if computations of states are run in parallel using graphical processing units. Since simulations are performed for 100 matrices per type of network architecture and the parameter space for global coupling strength W is traversed in 41 steps the computation time needs to be multiplied by 4100 to arrive at the time needed to simulate a single architecture type.

A

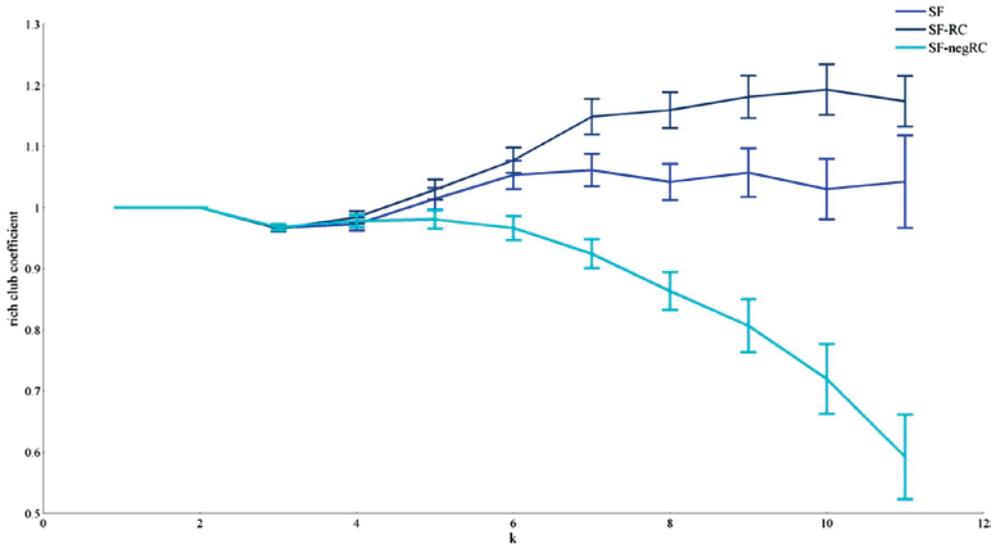


Figure A5.2: Normalized rich club coefficient (30 nodes). The figure shows the normalized rich club coefficient as a function of degree for the three types of scale free networks comprising 30 nodes. As observed for networks comprising 24 nodes, (figure 4.2, main text) SF-RC networks exhibit a mean rich club coefficient exceeding 1 in the interval [7,11] (dark blue line). SF networks whose hubs do not form a rich club exhibit a mean rich club coefficient around 1 (blue line); SF-negRC networks whose high-degree nodes show low levels of interconnectivity, exhibit a mean rich club coefficient consistently smaller than 1 (light blue line).

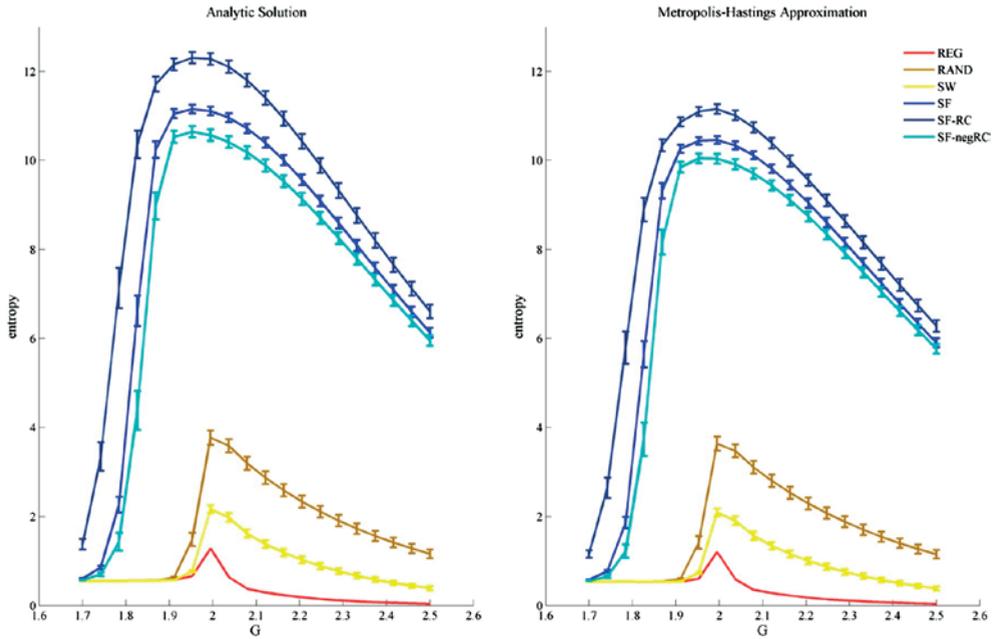


Figure A5.3: Entropy of the attractors of spin glass networks (30 nodes). The figure shows the entropy observed for each type of network architecture as a function of global coupling strength G . All networks contained thirty nodes and 180 edges. The left panel shows the semi-analytic solutions while the right panel shows approximations based on the application of the Metropolis-Hastings algorithm. In both cases, the scale free networks show larger entropy levels than the corresponding SW (yellow), REG (red) and RAND (orange) networks. Additionally, within the scale free class those exhibiting rich club organization (SF-RC, dark blue) show the largest levels of entropy, while those exhibiting negative rich club organization (SF-negRC, light blue) show the lowest levels of entropy. Again highly consistent with the results of the $N=24$, scale free networks without a rich club (SF, blue) show an intermediate level of entropy. The results of semi-analytic and approximate solutions are highly consistent and validate the use of the Metropolis-Hastings algorithm for estimation of relative effects with regard to global entropy in larger networks.

A

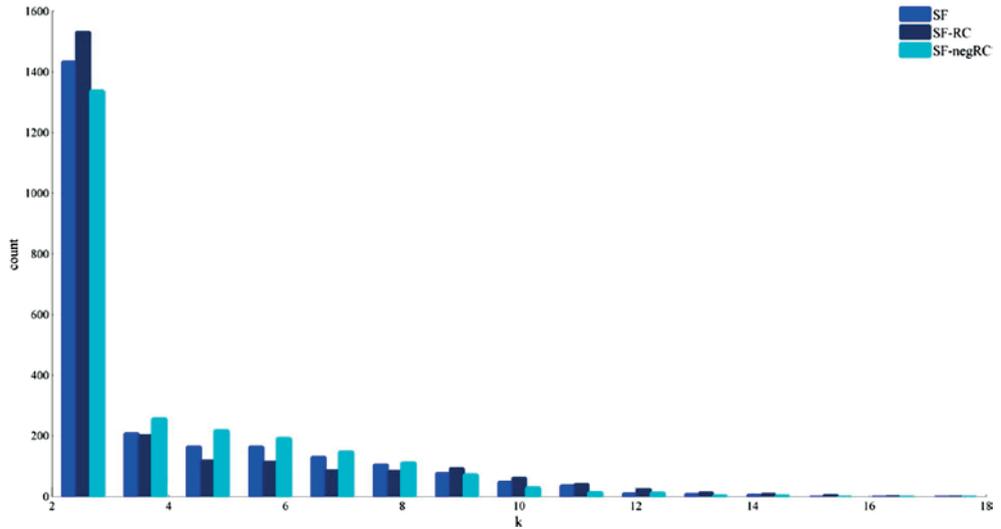


Figure A5.4: Degree distribution for scale free network architectures. The figure shows the number of nodes exhibiting a certain degree k over a range of degrees. As expected, low-degree nodes ($k = [2,5]$) are most abundant (76%) and high-degree nodes ($k = [11,14]$) least abundant (3%) for SF, SF-RC as well as SF-negRC. This overabundance of low-degree nodes implies that the diversity of the functional repertoire would greatly benefit if the low-degree nodes are functionally integrated. These effects are indeed observed in both scale free architectures containing a rich club and in empirical human networks (see main text and figures 4.5 & 4.8).

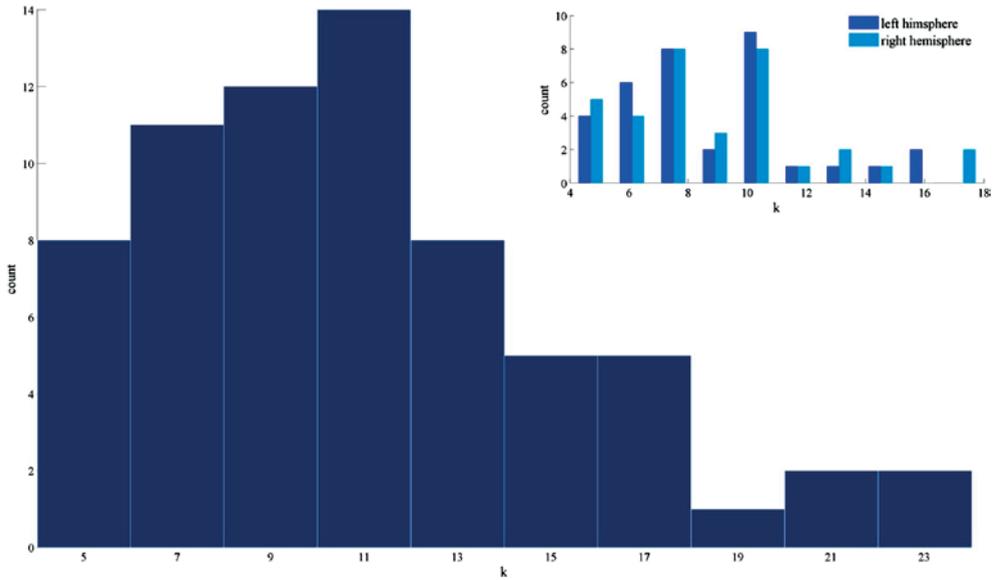


Figure A5.5: Degree distribution human data. The figure shows the number of nodes exhibiting a certain degree k over a range of degrees. The body of the figure shows the distribution for the complete human connectivity data comprising 68 nodes. Low-degree nodes ($k = [4,10]$) and moderate-degree nodes ($k = [11,17]$) together form the majority, constituting 46% and 43% of all nodes in the network, respectively. High-degree nodes ($k = [18,24]$) make up only a minority (12%). This overabundance of low and moderate-degree nodes implies that the diversity of the functional repertoire would greatly benefit if these nodes are functionally integrated. These effects are indeed observed. The inlay shows the intra-hemispheric degree distribution for the left and right hemisphere (each 34 nodes) separately. Low-degree nodes ($k = [4,8]$) and moderate-degree nodes ($k = [9,13]$) together form the majority, constituting 51% and 40% of all nodes in the network, respectively. High-degree nodes ($k = [14,18]$) make up only a minority (8%).

A

APPENDIX A6: SUPPLEMENTARY FIGURES CHAPTER 5

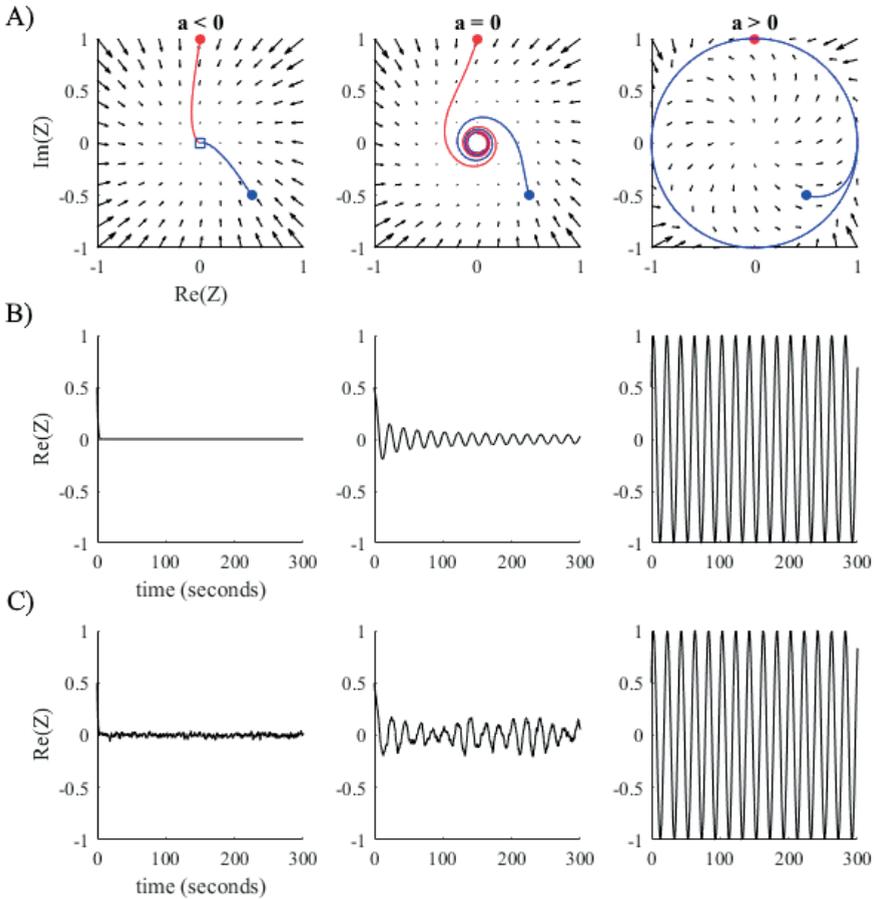


Figure A6.1: Local dynamics. This figure depicts the behavior individual (i.e. uncoupled) brain regions exhibit for different values of the local bifurcation parameter a . *Panel A* shows phase portraits detailing the trajectories of the complex variable z in terms of its real (x -axis) and imaginary (y -axis) components. For each value of a , two trajectories are highlighted. Circles represent initial conditions, while squares represent stable fixed points. The first trajectory (red) starts at initial condition $z = 0 + 1i$ while the second (blue) starts at $z = 0.5 - 0.5i$ for all values of a . If a is below zero a single stable (attracting) fixed point exists at $z = 0 + 0i = 0$. At $a = 0$ the system undergoes a bifurcation with a stable spiral towards the fixed point emerging. As a increases beyond zero, the fixed point becomes repellent and the system exhibits a stable limit cycle whose radius depends on the exact value of a . *Panel B* shows the behavior of the real component of z (i.e. what we treat as the simulated BOLD signal) corresponding to the three phase portraits presented in panel A. The model region had an intrinsic frequency of .05 Hz and was simulated for 300 seconds. If a is below zero the signal quickly decays. For $a = 0$ the signal oscillates at its intrinsic frequency with its amplitude decreasing over time. Finally, if a is larger than zero, the signal exhibits oscillations at fixed frequency and amplitude. *Panel C* shows again the simulated BOLD signal but in the presence of noise modeled as a Wiener process scaled by a factor $\beta = .02$. If a is below zero, the signal is essentially noise centered on zero. If a is equal to zero, the signal is a hybrid of noise and oscillations. If a is larger than zero, the signal is oscillatory.

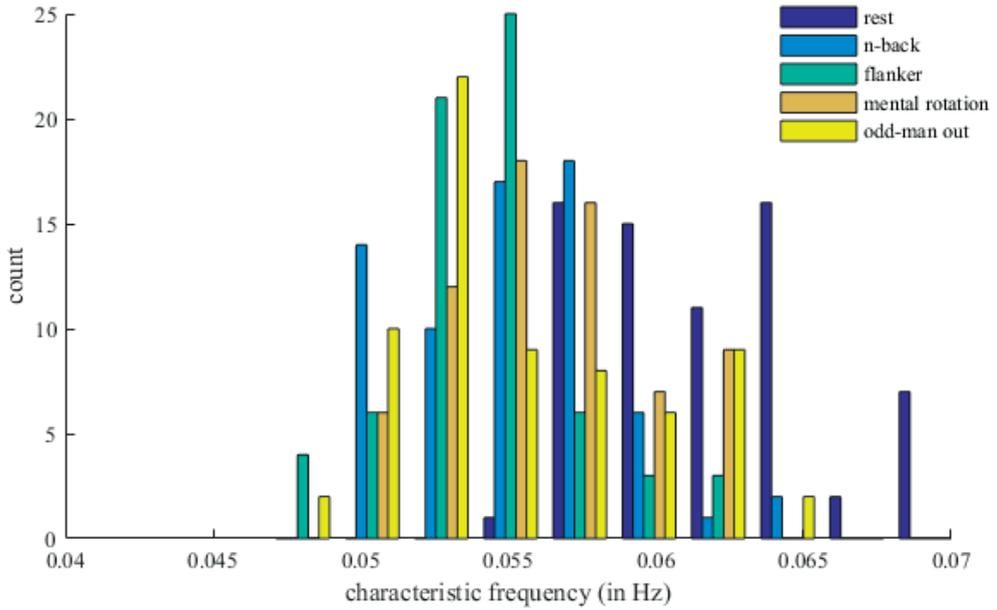


Figure A6.2: Characteristic Frequencies. This figure shows how distributions of characteristic frequencies across brain regions for each task. Distributions are largely similar for tasks but are shifted with respect to rest. According to a paired Kolmogorov-Smirnoff test, the distribution for rest differs from every task with all P-values being $\ll .001$. Among the tasks the mental rotation task differs significantly from both the flanker ($P = .004$) and odd-man out ($P = .039$) tasks.

A

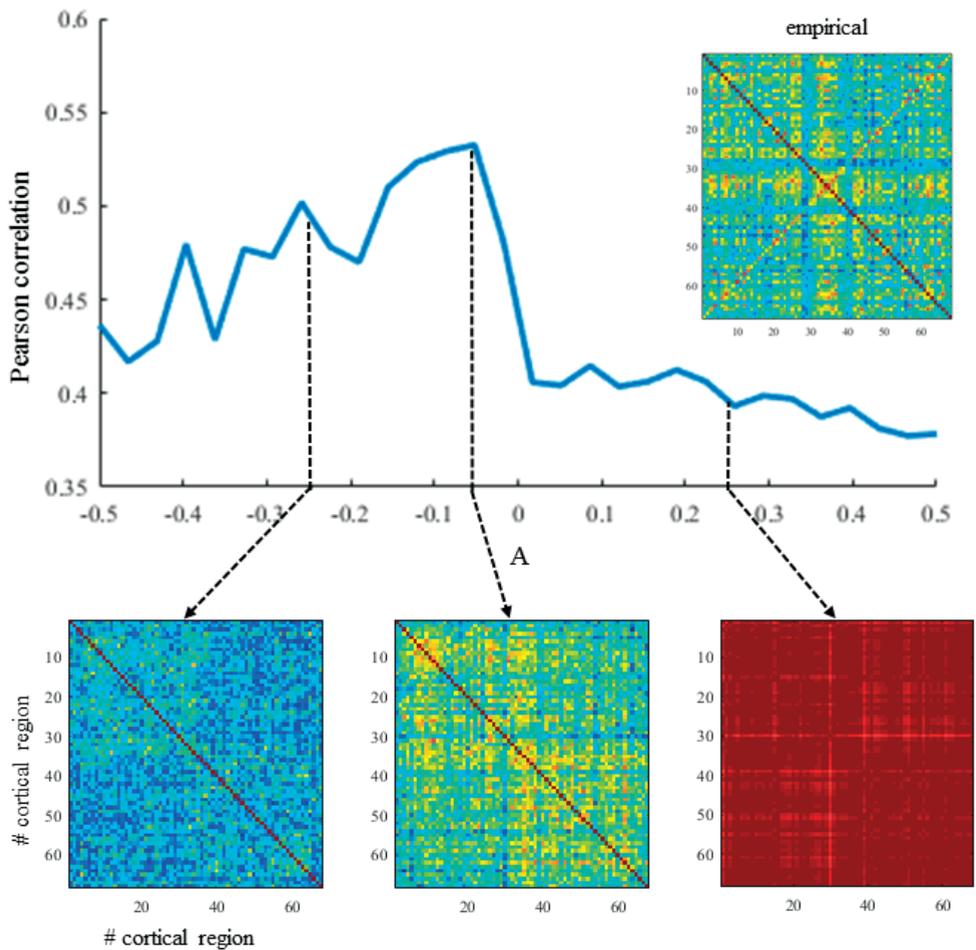


Figure A6.3: Functional Connectivity at Different Working Points. This figure shows how the Pearson correlation between simulated and empirical grand average functional connectivity observed during rest depends on the working point of the system. Specifically, Pearson correlation coefficients are depicted over the entire range of values explored for the bifurcation parameter A for a fixed value of global coupling ($G = .16$). The inlay on the top right corner shows the empirical functional connectivity the model needs to reproduce. The bottom row shows simulated functional connectivity observed for three exemplary values for A , namely: $A = -.25$, $A = -.06$ (optimal point), and $A = +.25$.

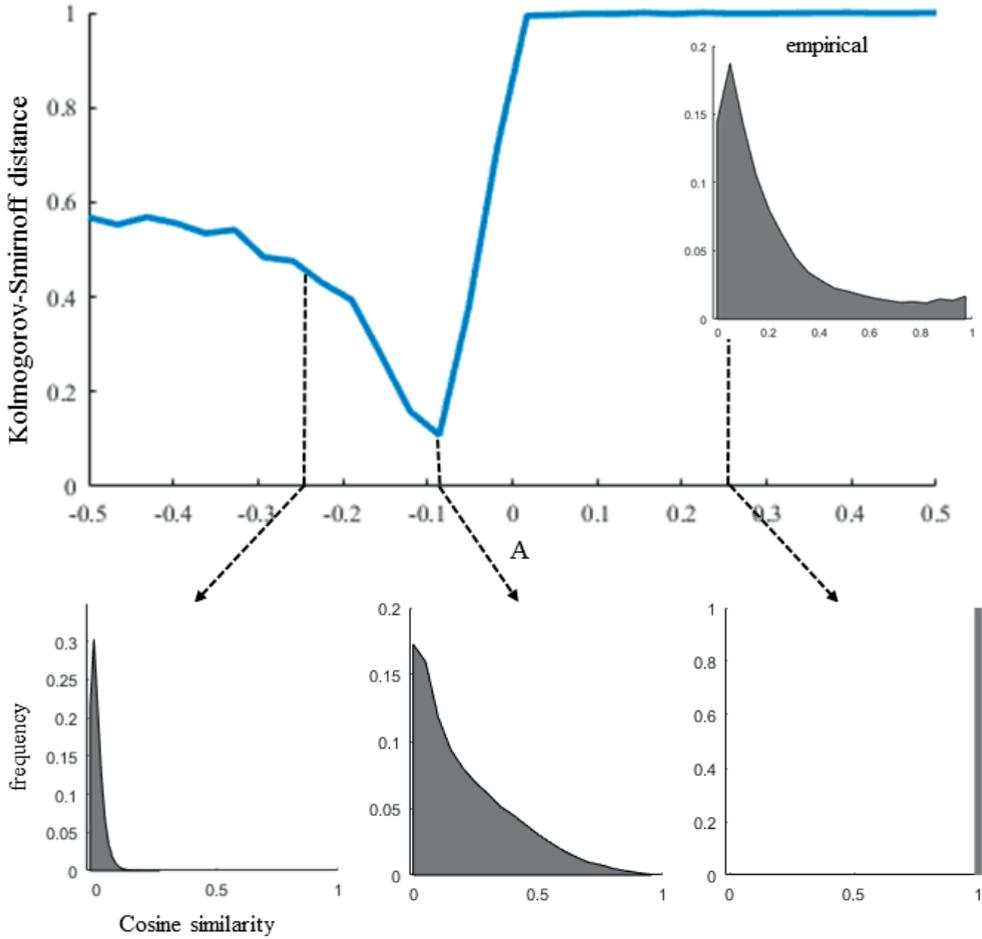


Figure A6.4: Dynamic Functional Connectivity at Different Working Points. This figure shows how the Kolmogorov-Smirnov distance between simulated and empirical dynamic functional connectivity observed during rest depends on the working point of the system. Specifically, Kolmogorov-Smirnov distances are depicted over the entire range of values explored for the bifurcation parameter A for a fixed value of global coupling ($G = .16$). The inlay on the top right corner shows the empirical distribution of dynamic functional connectivity (cosine similarity) values the model needs to reproduce. The bottom row shows simulated distributions observed for three exemplary values for A , namely: $A = -.25$, $A = -.085$ (optimal point), and $A = +.25$.

A

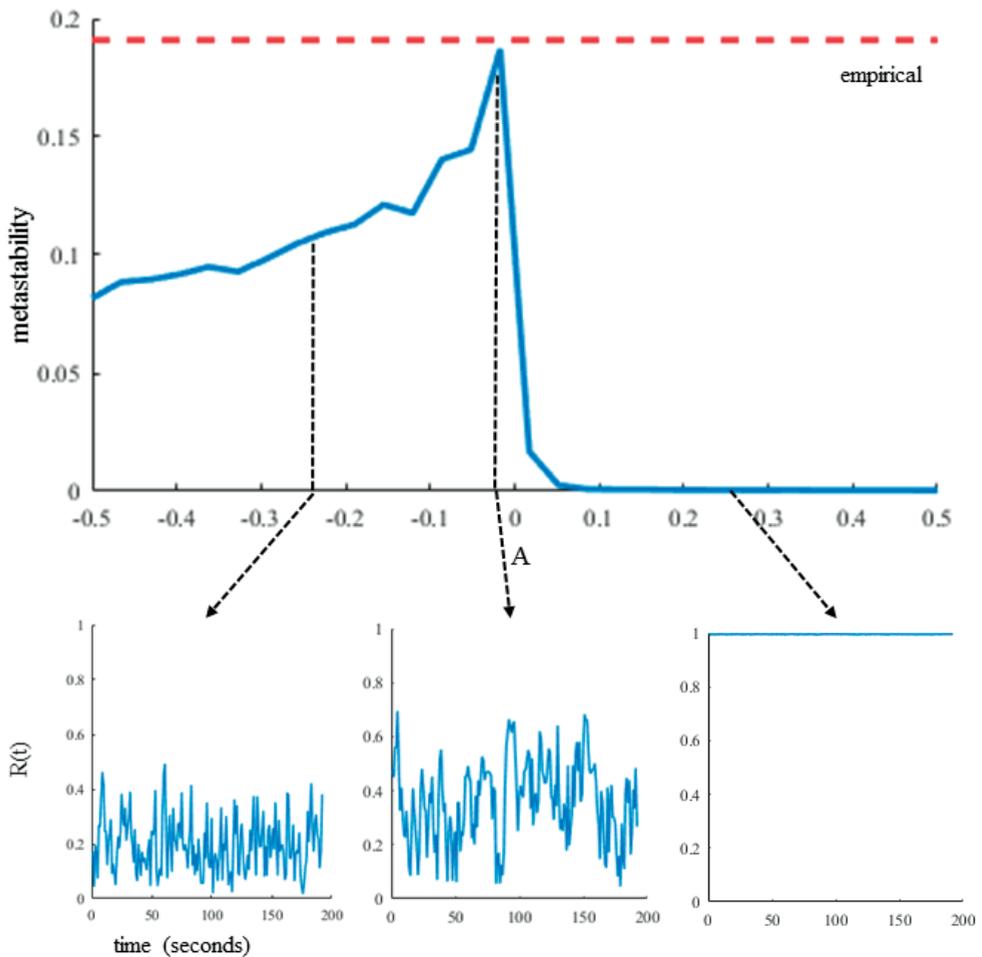


Figure A6.5: Metastability at Different Working Points. This figure shows how simulated metastability depends on the working point of the system. Specifically, metastability is depicted over the entire range of values explored for the bifurcation parameter A for a fixed value of global coupling ($G = .16$). The dashed red line indicates the average metastability observed for resting state data. The bottom row shows instantaneous Kuramoto order parameter values over the entire range of time produced by model simulations for three exemplary values for A , namely: $A = -.25$, $A = -.01$ (optimal point), and $A = +.25$.

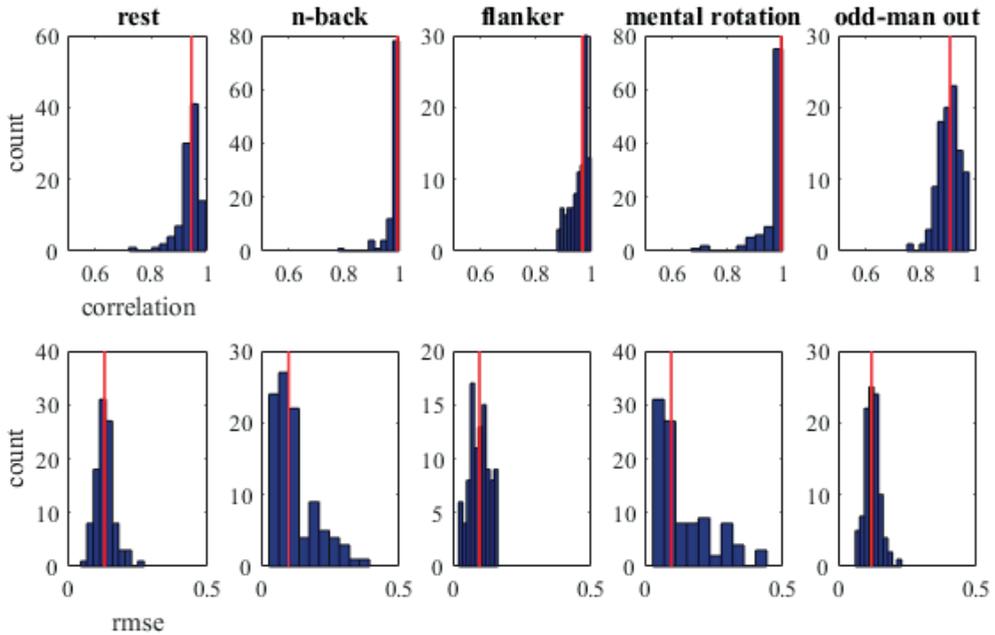


Figure A6.6: Convergence of local optimization process. *Panel A)* shows histograms of the correlations between local bifurcation parameters obtained from initializing each brain region with $a_j = -.0517$ and those obtained from 100 random initialization values per brain region. Median correlations were $r = .94$, $r = .99$, $r = .97$, $r = .99$, and $r = .90$ for rest and the four tasks, respectively. *Panel B)* shows histograms of the residual mean squared error (rmse) between local bifurcation parameters obtained from initializing each brain region with $a_j = -.0517$ and those obtained from 100 random initialization values per brain region. Median error rates were $rmse = .13$, $rmse = .10$, $rmse = .10$, $rmse = .10$, and $rmse = .12$ for rest and the four tasks, respectively.

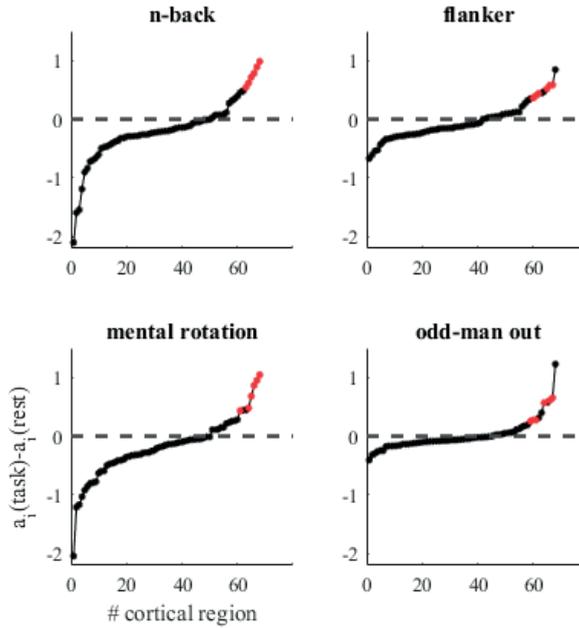


Figure A6.7: Ranking of cortical regions. This figure shows the ranking of cortical regions according to the differences in local bifurcation parameter between task and rest for all tasks. Rich club regions (marked in red) are among the 10 regions exhibiting the strongest differences in every task.

APPENDIX A7: ORDERING OF CORTICAL REGIONS

Table A7: 68 cortical regions

#	Region Name
1	left banks of the superior temporal sulcus
2	left caudal anterior-cingulate cortex
3	left caudal middle frontal gyrus
4	left cuneus
5	left entorhinal cortex
6	left fusiform gyrus
7	left inferior parietal cortex
8	left inferior temporal cortex
9	left isthmus-cingulate cortex
10	left lateral occipital cortex
11	left lateral orbital frontal cortex
12	left lingual gyrus
13	left medial orbital frontal cortex
14	left middle temporal gyrus
15	left parahippocampal gyrus
16	left paracentral lobule
17	left pars opercularis
18	left pars orbitalis
19	left pars triangularis
20	left pericalcarine cortex
21	left postcentral gyrus
22	left posterior-cingulate cortex
23	left precentral gyrus
24	left rostral anterior cingulate cortex
25	left rostral middle frontal gyrus
26	left superior temporal gyrus
27	left supramarginal gyrus
28	left frontal pole
29	left temporal pole
30	left transverse temporal cortex
31	left insula
32	left precuneus
33	left superior frontal cortex
34	left superior parietal cortex
35	right superior parietal cortex
36	right superior frontal cortex

Appendices

37	right precuneus
38	right insula
39	right transverse temporal cortex
40	right temporal pole
41	right frontal pole
42	right supramarginal gyrus
43	right superior temporal gyrus
44	right rostral middle frontal gyrus
45	right rostral anterior cingulate cortex
46	right precentral gyrus
47	right posterior-cingulate cortex
48	right postcentral gyrus
49	right pericalcarine cortex
50	right pars triangularis
51	right pars orbitalis
52	right pars opercularis
53	right paracentral lobule
54	right parahippocampal gyrus
55	right middle temporal gyrus
56	right medial orbital frontal cortex
57	right lingual gyrus
58	right lateral orbital frontal cortex
59	right lateral occipitalcortex
60	right isthmus–cingulate cortex
61	right inferior temporal cortex
62	right inferior parietal cortex
63	right fusiform gyrus
64	right entorhinal cortex
65	right cuneus
66	right caudal middle frontal gyrus
67	right caudal anterior-cingulate cortex
68	right banks of the superior temporal sulcus

VALORIZATION

Opinions differ on the question whether scientific knowledge is of instrumental or final value; that is, whether it is a means to an end or an end in and of itself. Supporters of the former view might hold that science being publicly funded ought to directly and tangibly benefit society. Opponents, on the other hand, deplore the degradation of scientific knowledge to the status of a mere resource and fear the rise of conflicts of interest. Between these diametrically opposed views exists a third according to which scientific knowledge is of contributory value. According to this view scientific knowledge is not a means to any specific end nor is it an end in itself, rather it contributes indirectly to an enlightened and free society.

I shall not discuss which view holds the most merit in my opinion nor will I discuss the work presented in the present thesis in terms of any of these three. Instead, I will argue that the work presented in this thesis is of importance to the neuroscientific community. If I am successful in that, my work will inherit its value from that of neuroscientific knowledge, irrespective of whether that knowledge turns out to be instrumental, final, or contributory. First, I will focus on the two parts of my thesis and discuss the functional value of my methodological considerations regarding theory testing and theory building. Following this, I will discuss the functional value of the results of each research chapter for the neuroscientific community.

The first part of my thesis was dedicated to demonstrating that functional magnetic resonance imaging (fMRI) can be used to compare competing theories instantiated as computational models. This is of importance to the neuroscientific community as it adds a vastly available, seminal, and non-invasive neuroimaging tool to pre-existing methods for evaluating computational models. Since crucial findings necessary to resolve long-standing conflicts between competing theories often come with the transition to novel methods, fMRI might tip the scale on a number of theoretical controversies among computational neuroscientists. The resolution of these controversies might come in the form of outright refutations of some theories in the light of others or, as in the present thesis, by revealing how competing theories can be integrated. In both cases neuroscience would benefit from the resulting insights as they lead to a deeper understanding of the brain and open new research avenues to pursue.

The second part of my thesis was dedicated to demonstrating that the integration of fMRI with computational modeling can be used to develop theories in neuroscience. In this case the value for the neuroscientific community directly stems from the complementary nature of the two fields. Theoretical considerations can inform empirical research as well as supplement analysis tools. In turn, new insights stemming from fMRI research can immediately be integrated into existing theoretical scaffolds and hence drive, evaluate, and constrain theory development. The latter is especially important as it prevents researchers in the field of computational neuroscience from getting carried away by abstractions of the brain which eventually bear little resemblance to the original target system.

The research on population receptive fields (pRFs) presented in the second chapter of my thesis is relevant for computational neuroimaging, a field which is becoming increasingly important in vision neuroscience. Since pRFs are at the heart of computational neuroimaging it was essential to evaluate their estimation procedures as well as establish their temporal consistency. Furthermore, knowledge of receptive fields is crucial for a number of neuroscientific applications: a) they provide a source of information for the reconstruction of stimuli from the blood oxygenation level dependent signal, b) they can serve as target for transcranial magnetic stimulation, c) they may assist function based alignment, d) they provide a spatial forward model for computational models, and e) they might give important insights with respect to theories of attention as well as into pathologies of visual cortex and brain development. In order for these applications to fully benefit from pRF estimation, it was essential to validate the estimation framework.

In line with the possibility that population receptive fields can be utilized as a spatial forward model for computational models, chapter 3 of this thesis presents new developments with regard to a common brain space. The common brain space is an idealized, retinotopically organized, surface representation of visual cortex which allows for the integration of empirical and simulated data in the same anatomical frame of reference. Such integration allows for the evaluation of computational models in terms of predictions they make with respect to the detailed spatial activation profiles in visual cortex. This is of importance for the neuroimaging community as it supplements existing techniques able to evaluate the information content of visually responsive brain regions predicted by computational models. That is, it enables neuroscientists to investigate not only what information is encoded in a region but also how that information is represented.

Another topic of great interest to the (neuro)scientific community pertains to the role of the rich club phenomenon present in the human connectome. This phenomenon is prominent in many network topologies from the internet to transportation networks and understanding its implications for the brain will aid us in understanding it in other topologies and vice versa. With regard to the brain, the fact that the rich club forms a central core in the connectome has lead researchers to suggest that it is important for the integration of information processing within specialized brain regions, and hence for cognition. However, studying the architecture of the connectome is insufficient to establish and further elaborate this hypothesis. It is necessary to study the functional implications of this architecture and contrast it with others. Chapter 4 presents simulation research addressing this issue by studying the implications different architectures, including those possessing a rich club, have for the capacity to functionally integrate distinct brain regions. As such the work I present there functions as a theoretical framework from which interesting empirical studies on the relationship between rich club organization and cognitive capabilities can be derived. Chapter 5 builds on this framework by introducing empirical findings and combining whole-brain simulations with fMRI. Of interest to the neuroscientific community is here especially that the work presented in chapter 5

provides further evidence for the thesis that the rich club is essential for functional integration as well as that it extends this thesis by suggesting a local mechanism through which rich club regions might help to organize and coordinate this integration. In accordance with what I set out to show here, the present work is thus relevant to the neuroscientific community in a number of respects.

ACKNOWLEDGEMENTS

At this point I would like to thank everyone who contributed to this thesis either directly or by virtue of their extended support. Each of you were indispensable for the completion of my PhD and I am deeply grateful. A special thanks I would like to extend to the members of the assessment committee and the corona for taking the time to evaluate my thesis and my defense.

First and foremost, I would like to thank my supervisors, Rainer Goebel, Gustavo Deco, and Judith Peters. Rainer, you have put such tremendous faith in me; always encouraging me to pursue my ideas. Whenever I started to doubt, your unrivaled enthusiasm never failed to invigorate me. You truly are a very unique supervisor and I am very glad to have had the opportunity to write my thesis under your guidance. I am also happy to remain within your group where I am sure you will continue to motivate and inspire me. Gus, I remember how nervous I was when composing my first email to you asking whether I could write my Master thesis in your lab. However, your response (yes sure) and the uncomplicated attitude you have displayed ever since quickly put me at ease. I learned an immense amount from you, not only with respect to science but also how to properly grill meat. I am glad I can continue to work with and learn from you. Judith, I have never met a person with a more positive attitude. Despite all your responsibilities and busy schedule your door was always open and you always greeted me with a smile. It was a great pleasure to work with you and I hope in the future we can pursue a few more projects together.

Next, I would like to extend my thanks to those with whom I had the opportunity to work and from whose experience I was able to learn. Joel, Sven, Valentin, and Bert, thank you for all the discussions over stimulus design, MRI setup, and experimental procedures. Martijn, Matthieu, Gorka, Niels, and Marcel, I feel very much enriched by our continued investigations of the rich club. Peter and Gesa, thank you for opening the world of psychophysics to me. I hope there will be more collaborations in the future! I would also like to thank all those whose instructions and lessons have set me on my path to science. Margarete Schäpers, Herbert Kremer, Rob de Vries, Herco Fontijn, Michael Capalbo, and foremost Arie van der Lugt.

I would like to thank the entire CN department for creating such a nice work atmosphere which is also in no small part due to my amazing office mates: Arne, Mehrdad, Thomas, Inge, Anke, and Anita. Furthermore, I would like to thank my fantastic paranymphs, Gojko and Arko, for supporting me during my defense.

While hard work is rightfully praised, it is often forgotten that leisure is equally of importance. Here, I would thus like to thank those colleagues and friends who made my leisure ever so rewarding. These last years would not have been as much fun, and indeed not as productive, if it was not for the extended coffee breaks and debates with Gojko, Jan, and Arne. Outside of the work environment, I look back with much pleasure at several burger sessions, New Year

celebrations, holidays, and get-togethers with Dietmar, Thorsten, Vita, Sabrina, Bianca, and Lia. You are the best!

Last but not least, I would like to thank my family who supported me over all these years. All of you have helped me remain in touch with reality outside the ivory tower of academia and have given me strength in difficult times.

My dearest parents, if it was not for your lifelong support and your hard work that created the necessary preconditions, none of this would have been possible. You have both been plagued by hardships but always refused to surrender without a fight. I owe everything to you.

Anna, it is hard to express how much you enrich my life. Your intellect challenges me, your tranquility calms me down, your silliness makes me laugh (even in my sleep), and your compassion makes me want to be a better person. Thank you so much

PUBLICATIONS

*Authors contributed equally to this work

Senden, M., Goebel, R., & Deco, G. (2012). Structural connectivity allows for multi-threading during rest: the structure of the cortex leads to efficient alternation between resting state exploratory behavior and default mode processing. *Neuroimage*, *60*(4), 2274–2284.

Senden, M., Deco, G., de Reus, M. A., Goebel, R., & van den Heuvel, M. P. (2014). Rich club organization supports a diverse set of functional network configurations. *Neuroimage*, *96*, 174–182.

Senden, M., Reithler, J., Gijssen, S., & Goebel, R. (2014). Evaluating population receptive field estimation frameworks in terms of robustness and reproducibility. *PLoS One*, *9*(12), e114054.

Senden, M., Reuter, N., Heuvel, M.P., Goebel, R.*; Deco, G*. (*in print*). Cortical rich club regions can organize state-dependent functional network formation by engaging in oscillatory behavior. *Neuroimage*

Senden, M., Peters, J., Jans, B., Reithler, J., & Goebel, R. (*in preparation*). Comparing computational models of simultaneous brightness contrast using ultra-high field functional magnetic resonance imaging.

Lange, G.*, **Senden, M.***, Radermacher, A., Goebel, R., & De Weerd, P. (*in preparation*).

Long term successive training causes behavioral interference

Reuter, N*., **Senden, M.***, Deco, G., Goebel, R.; Heuvel, M.P., Gilson, M. (*in preparation*). Whole-brain effective connectivity during resting and task states identifies cortical rich club as gated distributor.

Deco, G., **Senden, M.**, & Jirsa, V. (2012). How anatomy shapes dynamics: a semi-analytical study of the brain at rest by a simple spin model. *Frontiers in Computational Neuroscience*, *6*, 68.

Emmerling, T. C., **Senden, M.**, Frost, M. A., & Goebel, R. (*in preparation*). Reconstructing and decoding imagined letters from the visual system using ultra-high field fMRI.

Zimmermann, J., **Senden, M.**, De Martino, F., Goebel, R. (2014) Model free sub millimeter connective receptive field estimates at ultra high field strength. In Zimmermann, J., *The how, not where in functional magnetic resonance imaging* (doctoral dissertation).

INVITED TALKS

Senden, M. A combined neuroimaging and computational modeling approach to vision research. Dutch Neuro-Endo-Psycho meeting. May 31, 2012, Lunteren, the Netherlands.

Senden, M. Modeling local dynamics in a whole-brain network using the Jansen-Rit equations. Visit at Computational Neuroscience Group, Universitat Pompeu Fabra. September 28, 2012, Barcelona, Spain.

Senden, M. How anatomy shapes dynamics: Structural connectivity leads to a rich functional repertoire and an efficient exploration thereof. Donders Discussions. October 25, 2012, Nijmegen, the Netherlands.

Senden, M. Decision making & the brain. Lecture at Maastricht School of Economics. 2012 – 2014, Maastricht, the Netherlands

Senden, M. Rich club organization supports a diverse set of functional network configurations. Visit at Brain Center Rudolph Magnus. April 2, 2014, Utrecht, the Netherlands.

Senden, M. Bridging principles & reference spaces: How receptive fields link simulated with empirical data and provide a common reference space for analysis. Advanced ERC Grant “ColumnarCodeCracking” meeting. May 8, 2015, Salerno, Italy.

Senden, M. Unraveling intra-regional representations and computations by providing tailored predictions of voxel activations through population receptive fields. Visit at Netherlands Institute for Neuroscience, December 21, 2015, Amsterdam, the Netherlands.

Senden, M. High-level models of eye movement control. CDP4 kick-off meeting. June 14, 2016, Paris, France.

CURRICULUM VITAE

Mario Senden was born on the 5th of August 1985 in Aachen, Germany. He obtained his High School Diploma in 2005 from the Goethe Gymnasium in Stolberg, Germany. Subsequently, he performed his civilian service (Zivildienst) at the Psychiatric Hospital Alexianer Krankenhaus in Aachen, Germany. In 2006 Mario was admitted to the Psychology Bachelor program at Maastricht University, the Netherlands, which he finished in 2009 (cum laude). He was then admitted to the Cognitive Neuroscience Master program at Maastricht University from which he graduated in 2011 (cum laude). For his Master thesis, he investigated the effect of structural topology on network dynamics in a spiking neuron model of interconnected brain regions. This research he conducted under the supervision of Prof. Dr. Gustavo Deco at the Universitat Pompeu Fabra in Barcelona, Spain. For this work Mario received the student prize from the Stichting Wetenschapsbeoefening UM. From 2011 to 2016 Mario was a doctoral candidate at the Department of Cognitive Neuroscience at Maastricht University under the supervision of Prof. Dr. Rainer Goebel, Prof. Dr. Gustavo Deco, and Dr. Judith Peters. His research was on the integration of functional neuroimaging and computational modeling in the domains of early visual processing and structure-function relations in the brain. Mario currently works as a postdoctoral researcher within the Department of Cognitive Neuroscience investigating the computational mechanisms underlying visuomotor integration.