Large-scale functional MRI analysis to accumulate knowledge on brain functions

Yannick Schwartz

UNIVERSITÉ PARIS SUD

Comprendre le monde, construire l'avenir®

informatiques mathématiques

UNIVERSITÉ PARIS-SUD

DOCTORAL SCHOOL 427 : COMPUTER SCIENCE
PARIETAL TEAM - INRIA SACLAY

Large-scale functional MRI analysis to accumulate knowledge on brain functions.

Yannick Schwartz

A dissertation submitted for the degree of Doctor of Science,
in the subject of Computer Science.

Defended publicly the 21th of April 2015.

Advisors :	Dr Jean-Baptiste Poline	University California Berkeley, USA
	Dr Bertrand Thirion	Parietal Team - Inria Saclay, France
Reviewers :	Dr Finn Årup Nielsen	Technical University of Denmark, Denmark
	Pr Tor Wager	University of Colorado, USA
Examiners :	Dr Yves Burnod	Faculté de médecine Pierre et Marie Curie, France
	Pr Alain Denise	Université Paris-Sud, France
	Pr Simon Eickhoff	Heinrich-Heine University, Germany
	Dr Jean-Baptiste Poline	University California Berkeley, USA
	Dr Gaël Varoquaux	Parietal Team - Inria Saclay, France

UNIVERSITÉ PARIS SUD

Comprendre le monde, construire l'avenir®

informatiques mathématiques

Université Paris-Sud

Ecole Doctorale 427 : informatique Paris-Sud
Équipe Parietal - Inria Saclay

Discipline : informatique

Thèse de doctorat

Analyse à grande échelle d'IRM fonctionnelle pour accumuler la connaissance sur les fonctions cérébrales.

Yannick Schwartz

Présentée et soutenue publiquement le 21 avril 2015.

Directeur de thèse :	Dr Jean-Baptiste Poline	University California Berkeley, USA
Co-directeur de thèse :	Dr Bertrand Thirion	Parietal Team - Inria Saclay, France
Rapporteurs :	Dr Finn Årup Nielsen	Technical University of Denmark, Danemark
	Pr Tor Wager	University of Colorado, USA
Examinateurs :	Dr Yves Burnod	Faculté de médecine Pierre et Marie Curie, France
	Pr Alain Denise	Université Paris-Sud, France
	Pr Simon Eickhoff	Heinrich-Heine University, Allemagne
	Dr Jean-Baptiste Poline	University California Berkeley, USA
	Dr Gaël Varoquaux	Parietal Team - Inria Saclay, France

Abstract

How can we accumulate knowledge on brain functions? How can we leverage years of research in functional MRI to analyse finer-grained psychological constructs, and build a comprehensive model of the brain? Researchers usually rely on single studies to delineate brain regions recruited by mental processes. They relate their findings to previous works in an informal way by defining regions of interest from the literature. Meta-analysis approaches provide a more principled way to build upon the literature.

This thesis investigates three ways to assemble knowledge using activation maps from a large amount of studies. First, we present an approach that uses jointly two similar fMRI experiments, to better condition an analysis from a statistical standpoint. We show that it is a valuable data-driven alternative to traditional regions of interest analyses, but fails to provide a systematic way to relate studies, and thus does not permit to integrate knowledge on a large scale. Because of the difficulty to associate multiple studies, we resort to using a single dataset sampling a large number of stimuli for our second contribution. This method estimates functional networks associated with functional profiles, where the functional networks are interacting brain regions and the functional profiles are a weighted set of cognitive descriptors. This work successfully yields known brain networks and automatically associates meaningful descriptions. Its limitations lie in the unsupervised nature of this method, which is more difficult to validate, and the use of a single dataset. It however brings the notion of cognitive labels, which is central to our last contribution. Our last contribution presents a method that learns functional atlases by combining several datasets. [Henson 2006] shows that forward inference, i.e. the probability of an activation given a cognitive process, is often not sufficient to conclude on the engagement of brain regions for a cognitive process. Conversely, [Poldrack 2006] describes reverse inference as the probability of a cognitive process given an activation, but warns of a logical fallacy in concluding on such inference from evoked activity. Avoiding this issue requires to perform reverse inference with a large coverage of the cognitive space. We present a framework that uses a "meta-design" to describe many different tasks with a common vocabulary, and use forward and reverse inference in conjunction to outline functional networks that are consistently represented across the studies. We use a predictive model for reverse inference, and perform prediction on unseen studies to guarantee that we do not learn studies' idiosyncrasies. This final contribution permits to learn functional atlases, i.e. functional networks associated with a cognitive concept.

We explored different possibilities to jointly analyse multiple fMRI experiments. We have found that one of the main challenges is to be able to relate the experiments with one another. As a solution, we propose a common vocabulary to describe the tasks. [Henson 2006] advocates the use of forward and reverse inference in conjunction to associate cognitive functions to brain regions, which is only possible in the context of a large scale analysis to overcome the limitations of reverse inference. This framing of the problem therefore makes it possible to establish a large statistical model of the brain, and accumulate knowledge across functional neuroimaging studies.

Acknowledgements

I first would like to express my gratitude to my two advisors Jean-Baptiste Poline and Bertrand Thirion, who gave me this great opportunity and helped me along my PhD with their supervision. I also would like to thank Gaël Varoquaux, who was for me not only a shadow advisor, but also the perfect spoon for our various travels. Thank you for this amazing experience, and for introducing me to the realm of science. I am also grateful to Finn Årup Nielsen and Tor Wager, for taking time to review my thesis and all their insightful comments, as well as Yves Burnod, Simon Eickhoff, and Alain Denise for being part of my defense committee.

I would like to give special thanks to all the Parietal team, both present and past members: you are the reason why this team is so uniquely and trollingly disruptive. So big thanks to Vincent Michel (a.k.a. Nickel Mitchell) for showing me that technical skills were optional to do a PhD, and Alan Tucholka (a.k.a. the Graou) for teaching me the relational skills I needed in a professional environment, Cécilia Damon who was like a big sister to me, Alexandre Gramfort (malade), Pierre "Commissaire" Fillard, Philippe "tchoutchou" Ciuciu, Fabian "Chicken" Pedregosa, Viviana "No no no" Siless, Virgile Fritsch, petaflopic Benoît Da Mota, Michael Eickenberg, Olivier Grisel, Loïc Estève, Andrés Hoyos Idrobo, Alexandre Abraham, Konstantin Shmelkov, Danilo "triple that awsome" Bzdok, Aina Frau, Solveig Badillo, Salma Bougacha, Mehdi Rahim, Nicolas Chauffert, Ana Luísa Pinho, and Kamalakar Reddy. Special thanks to Régine Bricquet, Barbara Moulin, and Maike Gilliot who saved me from many administrative hurdles.

I have spent a bit more than seven and a half years at Neurospin, and had the chance to meet many people that made this experience rich both on the professional and personal levels. It may just look like a long enumeration of names but I really enjoyed working and/or spending (too much) time around a coffee with each and every single one of you. Thanks to all the IMAGEN team, in particular Benjamin Thyreau, who taught me right from wrong, Alexis Barbot, Fanny Gollier-Briant, and Eric Artiges. I would also like to acknowledge my adoptive parents Edouard Duchesnay and Dimitri Papadopoulos, as well as my collaborators from the other teams: Julien Lefèvre, David Germanaud, Clara Fischer, Pauline Roca, Edith Le Floch, Soizic Laguitton, Matthieu Perrot, Pamela Guevara, Urielle Thoprakarn, Grégory Operto, Mathieu Dubois, Yann Cointepas, Denis Rivière, Alexis Roche, Laurent Risser, Nicolas Souedet, Jean-François Mangin, Vincent Frouin, Mélanie Strauss, Denis Engemann, Murielle Fabre, Laura Dupas, Elodie Cauvet, Catherine Wacongne, Marie Amalric, Valentina Borghe-

sani, Christophe Pallier, Lucie Charles, Anne Kozem, and Aurélien Massire. And I also would like to congratulate Baptiste Gauthier for his PhD.

Thanks also to all my friends who reminded me that there is more to life than academic research: my forever roommates Colin and Rémi, Ahmed, Pieb, Piouf, Aurélien, Ienien, Chris, and Soubouzou San. Finally special thanks to all my family, in particular to my parents, and Cédric and Christel for their continued support. Last but not least, thanks to my beloved Rebecca, who whipped me when I needed, and believed in me when I could not.

Contents

Introduction

Context

Understanding how the human brain is the physical basis of the mind is one of the greatest challenges of this century. From synapses –connection between neurons– to the mind, the study of the brain has given rise to many scientific fields that examine this organ from the biological to the psychological angle. *Cognitive neuroscience* aims to understand how the neural substrates lead to the emergence of thoughts, and relies on several tools at its disposal including: lesion studies, which relate brain injuries to cognitive impairments, and *neuroimaging*, which uses brain imaging to test the validity of cognitive models.

Functional magnetic resonance imaging (fMRI) is an imaging technique that measures the level of oxygenation in the blood over time, which is linked to the neural activity. It is one of the most widely used imaging technique in cognitive neuroscience, as it provides an indirect but non-invasive access to the *neural code*, and enjoy a good spatial resolution. The most widespread experimental procedure to localize brain functions with fMRI, is to manipulate tasks –composed of several mental processes– in order to isolate a single mental process and assess how it affects brain activity. Standard analyses rely on *statistical inference* to produce *statistical parametric maps* or *activation maps* to evaluate brain activation in respect to a cognitive process. But each individual fMRI study only sheds light on a small fraction of the *cognitive space*, as only a limited number of stimuli may be presented in a single imaging session. One approach to overcome these limitations are *meta-analysis* methods.

Quantitative meta-analyses are an ensemble of methods that may be used with different goals in mind. The orginal use is to increase the statistical power on a given cognitive question by aggregating data from multiple studies. It is also used to summarize the vast and ever growing fMRI literature. This integration enables for example to give additional insights on the function of particular brain structures, by combining studies that used different experimental settings. Most fMRI meta-analyses rely on summarized data: the coordinates of the activation peaks from experiments. Some methods use the full statistical maps to perform meta-analyses but are not as common due to the difficulty of gathering enough fMRI experiments. Overall, there has been so far little incentive to share neuroimaging data partly because few methods are available to take advantage of an accumulation of data. This is a known chicken and egg problem, but vast amounts of available data opens new opportunities.

One such opportunity is to relate cognitive processes from different fields between them, and provide a more comprehensive view of the brain. Importantly, combining datasets enables to go beyond individual studies, which only reflect the cognitive theories being tested. In this thesis we investigate the use of reverse inference, which reasons backwards from brain activity to draw conclusions on a cognitive process instead of a particular task. This is a known logical fallacy: researchers are tempted to make *informal* reverse inferences by interpreting brain activity in respect to the literature, and concluding that a region is specific to the cognitive process under study. A large amount of data that spans a large cognitive coverage however opens the possibility to perform reverse inference in a principled way. This approach, when conducted on a large-scale, is particularly promising as it enables to associate sets of regions with sets of cognitive processes.

The overall objective of this thesis is to develop tools to jointly analyse multiple functional MRI studies. Outcomes of this thesis include the learning of functional networks, and atlases of brain functions.

Organization and contributions of this thesis

Part I - State of the art: a brief introduction to neuroimaging

Chapter 1 - From brain images to the sudy of the mind

This first chapter introduces functional imaging techniques and how they can be used to localize brain functions. More specifically, we summarize a few of the assumptions underlying the experimental setup in cognitive neuroscience, and how these experiments are used to complement the knowledge gained by brain lesion analyses. We show that functional MRI remains today one of the best non-invasive tools to map functions to structures in the brain.

Chapter 2 - Tools for neuroimaging data modeling

In the second chapter, we describe the statistical methods commonly used to analyse fMRI data, and the different meta-analysis approaches used by the community to analyse jointly multiple studies. We present in particular the statistical inference framework used by the majority of individual fMRI experiments to model the fMRI BOLD signal, and how it produces *statistical maps* that represent brain activity. We also introduce statistical learning methods for the particular context of neuroimaging. We finish by summarizing the meta-analyses methods for fMRI, specifically we describe coordinate-based meta-analysis and image-based meta-analysis methods, and how they are used to overcome some of the limitations of individual studies.

Part II - Contributions: from an image database to learning brain functions

Chapter 3 - Scaling up from individual studies

In this chapter, we assemble a large fMRI database by accumulating images from multiple individual studies. We review different datasharing initiatives

from which we pool our data, and how we organize the data. At a large-scale, data organization is key to automate data integration and processing streams. We describe our strategy to automate the fMRI pre-processing, statisical modeling, and data curation, and discuss the main bottlenecks and possible solutions to set up a large database. We finish by giving an overview of the accumulated data.

Chapter 4 - Functional localization by meta-analysis

Chapter 4 is our first attempt to combine multiple studies. We draw inspiration from region of interest (ROI) based analyses, that define ROIs from the literature or external datasets to better condition an analysis, but do so in a more principled way. We describe two contributions in which we use pairs of tasks in fMRI experiments to define common regions of interest. The first contribution examines the ability of one task to better condition testing for a second similar task. The major drawback of this method is that it relies on a manual threshold to define the regions of interest. The second contribution provides a methodology that aims to solve this shortcoming: it uses the ability of a classifier to generalize from one task to another as a test to select the threshold at which we define our regions of interest. This contribution can be seen as a multivariate alternative to contrasts conjunction. The main limitation of these methods is their poor scalability: they rely on the manual selection of fMRI experiments to compare, and do not give a broad overview of brain functions.

Chapter 5 - Learning functional networks

This chapter takes a step back from multi-study analyses. We previously focused on regions of interest but more modern approaches view the brain functioning within a set of distributed networks. The study of functional networks mainly relies on the use of unsupervised decomposition methods on resting state fMRI. These approaches however do not enable to automatically associate brain maps with functional labels. The contribution described in this chapter proposes an alternative way to expose functional networks on task fMRI, and associates them with what we call functional profiles. We rely on an unsupervised approach, that learns jointly a dictionary of functional profiles and a set of spatial maps. fMRI studies outline mental processes by combining experimental conditions. The functional profiles in this chapter reflect the response magnitude of the experimental conditions for each network, and enable to define a functional signature.

Chapter 6 - Learning functional atlases

This last chapter presents a supervised framework that benefits from a growing accumulation of datasets to associate brain networks with cognitive functions. We investigate the use of cognitive ontologies to define a *meta-design* that enables to co-analyse multiple task fMRI datasets despite the lack of common paradigm. Our framework uses forward and reverse inference in conjunction to map functional labels to the brain. Forward inference is commonly used to associate a brain region with a cognitive theory implemented by a dissociation of two experimental conditions. The two

conditions test for an effect of interest under an experimental paradigm, but do not guarantee that the resulting region is specific to that effect. A large-scale setting opens the possibility to invert the statistical inference and to reason from the brain activations to the mental processes. The combination of the two inferences gives more evidence that a cognitive process can be associated with a particular brain network, and enables to establish functional atlases.

Part I

State of the art: a brief introduction to neuroimaging

1 From brain images to the study of the mind

In this thesis, we investigate methods to combine multiple functional imaging studies, to provide a more comprehensive view of the brain functions. In this first chapter, we introduce functional imaging techniques and how they can be used to localize brain functions. More specifically, we summarize a few of the assumptions underlying the experimental setup in cognitive neuroscience, and how these experiments are used to complement the knowledge gained by brain lesion analyses. We finish by presenting the main functional areas in the brain.

Contents

1.1 From brain lesions to functional imaging

In 1861, the French physician Paul Broca heard of a patient from the Bicêtre Hospital suffering from a loss of speech but not a loss of comprehension nor other cognitive functions. This patient nicknamed "Tan" after the only word he could articulate intelligibly, died a few days later which enabled Broca to perform an autopsy. He discovered a lesion in the frontal lobe of the left cerebral hemisphere, next to the lateral sulcus. He confirmed his findings over the next years with autopsy evidence from additional patients. Patients suffering from this type of lesions are categorized as having *expressive aphasia*, which involves impairments specific to language production. This research enabled Broca to associate a brain region now known as *Broca's area* with a cognitive function: speech production. The kind of reasoning that identifies the neural substrate of a particular brain function by demonstrating that a lesion in a brain structure alters function X but not function Y is called a *single dissociation*. Single dissociation inferences indicate any of the following: i) the two functions have a different neural substrate, ii) the two functions are part of the same system but the damaged one is downstream from the preservered one, or iii) the maintained function requires fewer cognitive resources than the other one.

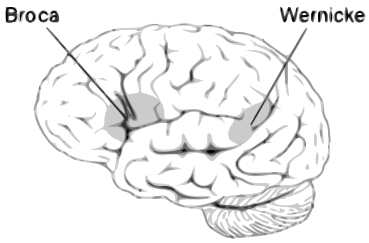


Figure 1.1: Broca's and Wernicke's areas were linked to two distinct language disorders in the late 19^{th} century, respectively expressive and receptive aphasia. Source: adapated from `www.wikipedia.org`

Shortly after Paul Broca's discovery, German physician Karl Wernicke made a converse observation, where patients where suffering from language comprehension disorders but not speech production. Using a similar procedure, he uncovered a region in the posterior section of the left superior temporal gyrus that is now referred to as *Wernicke's area*. The same single dissociation reasoning indicates that this region is involved in the understanding of written and spoken language. Damage in this region causes a disorder known a *receptive aphasia*. These two findings taken jointly are a case of *double dissociation*. Double dissociations on brain functions are possible when a lesion on a brain region A affects function X but not Y, and a lesion on a separate region B affects function Y but not X. Double dissociations yield much stronger conclusions than single dissociations, and demonstrate that two functions are separate, and have a different neural substrate. Even though recent research gives a more complex view of language processing, this seminal double dissociation paved the way for modern neuropsychology.

These kind of approaches are however limited to patients having lesions in an area of interest to the researcher. The localization of lesions could in addition only be done post mortem, which considerably slowed down the research process. The advent of brain imaging in the 1920s triggered a more systematic investigation of the brain structures and functions.

1.2 BOLD Functional MRI

Several brain imaging techniques exist today, and enable to study the brain structures, as well as its functions. *Anatomical imaging* focuses on the brain structures and tissues, whereas *functional imaging* detects physiological changes that indicate brain activity. This section presents BOLD functional imaging (fMRI), that we use in this thesis. fMRI has a good spatial reso-

lution ($1 - 3mm^3$) and a temporal resolution on the order of 1-3 seconds. MRI images are actually stacks of slice images, each of which is acquired in about 40-60 miliseconds, and have a full brain coverage. These properties make this imaging modality a good candidate for the *localization* of brain functions. Other functional imaging techniques such as Electroencephalography (EEG) or Magnetoencephalography (MEG) have a poor spatial resolution, but enjoy a temporal resolution in the order of the milisecond, which makes it possible to study the temporal dynamics of the information processing in the brain. Imaging modalities are generally a tradeoff between spatial and temporal resolutions, and are therefore tailored to investigate certain aspects of the brain functions.

Blood oxygenation level-dependent contrast Oxygen transportation in the organism is done through the circulatory system, and in particular with hemoglobin, a protein present in blood cells that has the ability to bind oxygen. Hemoglobin is found in two forms: oxyhemoglobin when it is oxygenated, and deoxyhemoglobin when it is deoxygenated. When oxyhemoglobin releases its oxygen atoms to turn to deoxyhemoglobin, it causes a difference in magnetic susceptibility. This contrast is known as blood oxygenation level-dependent (*BOLD*), because of its dependence on the level of oxygenation of the blood. This discovery by Ogawa *et al.* in 1990, has enabled to observe the *BOLD* contrast through a gradient-echo EPI (EchoPlanar Imaging) sequence, and introduced MRI as a functional imaging tool [14, 12] (see Figure 1.2)

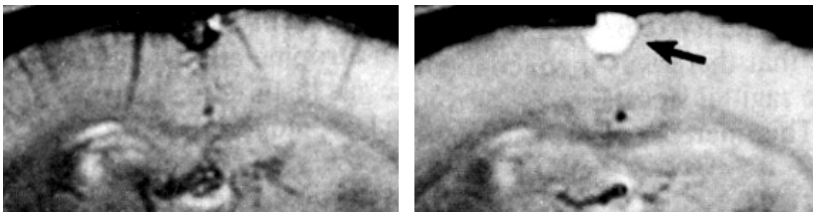


Figure 1.2: Ogawa et al. [12] original experiment on rats to bring to light the *BOLD* contrast: Left - Coronal slice showing the *BOLD* contrast of an anesthetized rat which has breathed pure O_2. Right - Coronal slice of the same rat, showing the *BOLD* contrast after respiration of a mixture of 90% of O_2 and 10% of CO_2 (this mixture increases the oxygenation of the venous blood). The arrow shows the sagittal sinus, which is one of the major veins of the brain. This picture shows a strong increase of intensity in this vein, and illustrates that the variation of blood oxygenation is visible in *BOLD* contrast.

The hemodynamic response At rest, the brain consumes the oxygen transported by the hemoglobin, and turns oxyhemoglobin into deoxyhemoglobin. Performing a cognitive task causes an oversupply of the oxygenated blood in the active region, and results in a decrease of deoxyhemoglobin. The oversupply process is called the *hemodynamic response*, and enables to relate the oxygenation level to neural activity [3, 6]. The decrease of deoxyhemoglobin implies a local increase in the average blood oxygenation, which can be observed with a *BOLD* contrast. As a consequence, *BOLD* captures changes in hemodynamics, that in turn are associated with neural activity, and makes it possible to use *BOLD* for functional imaging. Further work from Ogawa et al. [13] shows that a visual stimulation increases the intensity of the *BOLD* signal in the visual cortex. The use of MRI to detect *BOLD* is referred to as functional MRI (fMRI), or BOLD fMRI.

The *BOLD* response occurs with a delay after a stimulation, and does not return to baseline immediately. Typically the *BOLD* signal reaches its

maximum 5 seconds after the stimulation, and takes as long as 30 seconds to return to the baseline with an undershoot. The *Hemodynamic Response Function* (HRF) models this response pattern. The slow dynamic of the HRF explains the poor temporal resolution of fMRI. The HRF differs across brain locations and individuals [1, 8], but we generally consider a canonical HRF. Glover [7] proposed to define the HRF as a sum of two gamma functions, where the first gamma function models the initial stimulus response and the second gamma function models the undershoot:

$$h(t) = \frac{t^{\alpha_1 - 1}\beta_1^{\alpha_1} e^{-\beta_1 t}}{\Gamma(\alpha_1)} - c\frac{t^{\alpha_2 - 1}\beta_2^{\alpha_2} e^{-\beta_2 t}}{\Gamma(\alpha_2)} \tag{1.1}$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2$ control the shape and and scale, respectively, and c determines the ratio of the response to undershoot. Glover [7] estimated two HRFs, one in an auditory task and one in a motor task. Figure 1.3 shows the HRF based on the parameters estimated from the auditory task.

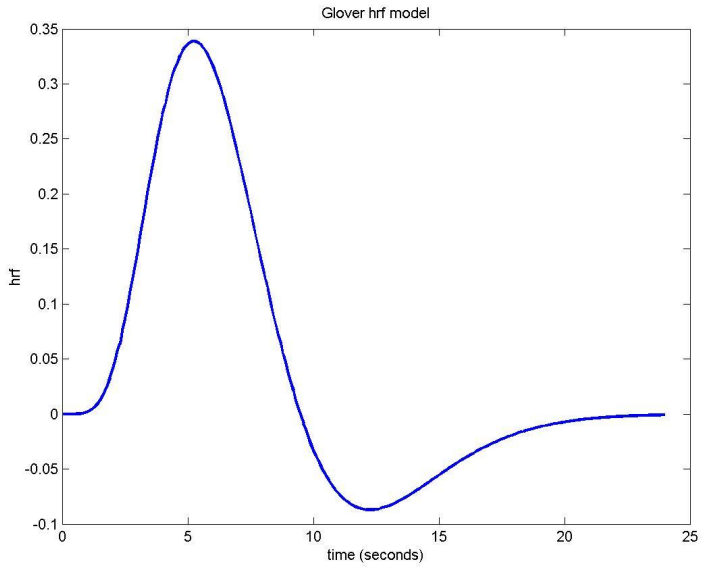


Figure 1.3: The Hemodynamic Response Function (HRF) first follows an increase of the signal ($1 - 5.2s$), then decreases ($5.2 - 12.2s$), and finally returns to baseline with undershoot ($12.2 - 20s$). Source: `http://www.math.mcgill.ca/keith/BICstat`

The BOLD signal is noisy, as it is impacted by various processes that are not related to cognition such as respiration and heart beat, as well as acquisition artifacts such as scanner drifts and noise distortions.

1.3 Mapping mental processes to the brain

The ability of fMRI to localize active regions during a task, makes it a powerful measurement technique for cognitive neuroscientists. Cognitive neuroscience studies the neural substrates of mental processes, and in particular how the cognitive processes are produced by the underlying brain circuitry. The mental processes refer to all kind of cognitive functions, which include perception, memory, speech production, decision making, emotion, motor, etc. fMRI gives an indirect access to the neural substrate, and as a consequence cognitive neuroscience has to rely on different assumptions to relate it to the mind. This section describes some of the assumptions and exper-

imental settings in cognitive neuroscience, and gives an overview of the main functional areas.

1.3.1 The cognitive neuroscience setting

The core assumption many fMRI studies is the **functional specialization**, or functional specificity. It means that we can expect to localize into specific and distinct brain regions, a large amount of the cognitive functions. Studies of the motor cortex is a good example that gives credit to this view, as reports indicate that fMRI activation is able to provide a clear somatotopic representation (see Figure 1.4) of foot, elbow, fist, thumb, index finger, and lip movements [9].

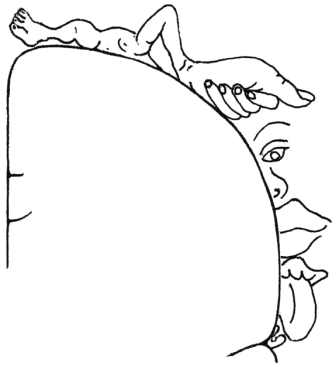


Figure 1.4: A motor homunculus which locates body parts along the postcentral gyrus. Source: `http://en.wikipedia.org/wiki/Somatotopic_arrangement`

Neuroimaging researchers that manipulate a task assume that is it possible to decompose this task into specific processes that can be manipulated independently. The **decomposability of mental processes** is the assumption underlying the experimental designs that explores how manipulations affect brain activity, in order to map neural and mental processes. The brain to mind mapping holds with the assumption that the decomposition is valid. This assumption has its critiques [17, 18] and supports [2], but provides the basis for several experimental designs.

Subtraction method One of the most commonly used experimental design is neuroimaging is the *subtraction method*. This method considers two experimental conditions which are supposed to only differ by a single cognitive process (Figure 1.5). Contrasting both conditions enables to *isolate* the cognitive process. This relies on the *pure insertion* assumption, which means that a single cognitive process can be inserted into another set of cognitive processes without affecting the rest. There are different issues with the pure insertion assumption, one of which is the problem of behavioral confounds: a perfect control for a given condition may never be found as it is often less demanding than the task of interest. This causes interpretation problems [5] since differences in brain activity may as well be imputed to the cognitive process under study, as to a difference in attention for instance. The subtraction method is still very common because of its simplicity, and researchers can use multiple controls with varying difficulties to mitigate the undesired confounds.

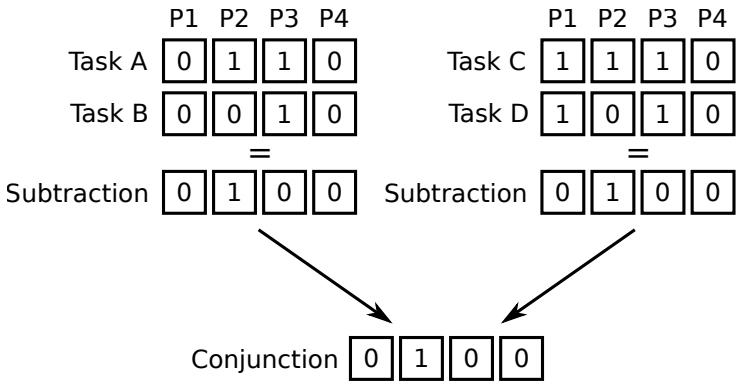


Figure 1.5: Subtraction design and conjunction, where $P1 \ldots P4$ represent the cognitive processes of the tasks. Source: Poldrack [15]

Cognitive conjunction Price and Friston [16] introduced an experimental design called *cognitive conjunction* (Figure 1.5) to avoid relying on *pure insertion*. This design uses two subtraction tasks, where only a single cognitive process is shared across the subtractions. A significant conjunction requires that each subtraction shows an activation, and that they do not differ in a significant manner. The revision of conjunction by Friston et al. [4] only identifies regions commonly active at a given threshold across the subtractions. Finally Nichols et al. [11] gives a valid approach for conjunction from the statistical standpoint. The main problem of these approaches is that they require all the subtractions to have a single cognitive process in common, which is objectively hard to assess. They also are difficult to perform on noisy data such as fMRI, as they have a lower sensitivity (*i.e.* lower detection power) than individual subtractions.

Factorial design *Factorial designs* require the pure insertion assumption, but manipulate multiple experimental factors at the same time. This kind of design use an analysis of variance to assess the main effect, as well as the interactions that occur when the effect of one factor varies depending on the manipulation of another variable. For example, a factorial design on face recognition may use *faces* and *houses* stimuli in different *sizes* or *orientations* as experimental conditions, to measure the main effect of *face* and the interactions incurred by *size* or *orientation*. Figure 1.6 illustrates a factorial design, where $P1 \dots P4$ represents the main effects (*e.g.* faces) and $C1 \dots C4$ a modulating variable (*e.g.* orientation). This method may therefore give additional evidence to map cognitive and neural processes.

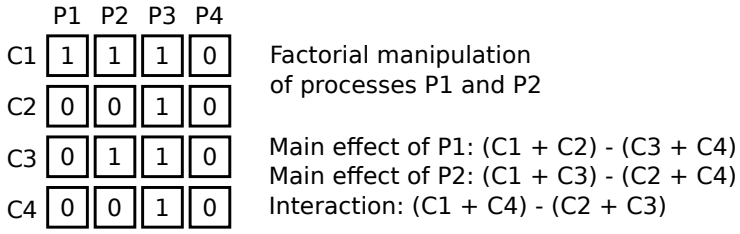


Figure 1.6: Factorial design, where $P1 \dots P4$ represent the cognitive processes of the tasks, and $C1 \dots C4$ the modulating variables. Source: Poldrack [15]

1.3.2 Main functional areas

Functional imaging techniques in pair with experimental designs enable to assign cognitive functions to brain structures. We present in Figure 1.7 the main functional and anatomical areas. The functional regions can be broadly categorized into three categories: sensory areas (*e.g. visual cortex, auditory cortex*) that process sensory inputs, motors areas (*e.g. primary motor cortex, premotor cortex*) that controls body parts, and associative areas (*e.g. Broca's area, prefrontal cortex*) that integrate higher-level information related to cognition [10]. This thesis will investigate a fair amount of these cognitive functions, as we aim to develop methods that yield functional atlases of the brain from the combination of multiple studies.

1.4 Conclusion

This chapter has introduced functional imaging techniques, in particular functional MRI, and how it can be used to access the neural code. We have also presented experimental paradigms that are commonly used by cognitive scientists to manipulate tasks and isolate mental processes. The combination of these two approaches have given neuroimaging researchers an increadible opportunity to map mental to neural processes without resorting to invasive methods, or post mortem analyses. Despite limitations of fMRI –especially regarding its temporal resolution– and the controversies that surround the underlying assumptions of experimental settings, fMRI remains as of today the best non-invasive tool to localize functions in the brain.

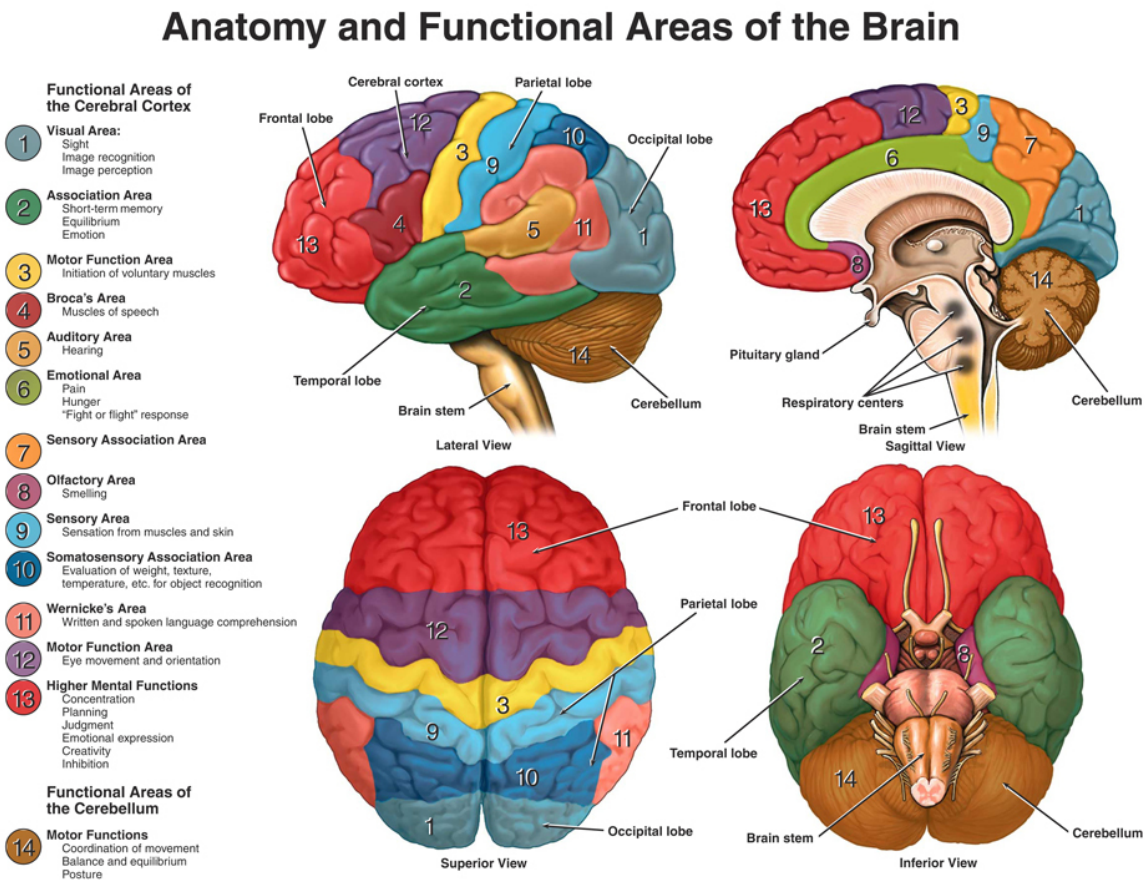


Figure 1.7: Main functional and anatomical areas of the brain. Credit: Nucleus Medical Art, Inc./Getty Images

Bibliography

[1] Solveig Badillo, Thomas Vincent, and Philippe Ciuciu. Group-level impacts of within-and between-subject hemodynamic variability in fMRI. *Neuroimage*, 82:433–448, 2013.

[2] William Bechtel. Decomposing the mind-brain: A long-term pursuit. *Brain and Mind*, 3(2):229–242, 2002.

[3] Geoffrey M Boynton, Stephen A Engel, Gary H Glover, and David J Heeger. Linear systems analysis of functional magnetic resonance imaging in human v1. *The journal of neuroscience*, 16(13):4207–4221, 1996.

[4] Karl J Friston, Andrew P Holmes, CJ Price, C Büchel, and KJ Worsley. Multisubject fMRI studies and conjunction analyses. *Neuroimage*, 10(4):385–396, 1999.

[5] KJ Friston, CJ Price, Paul Fletcher, C Moore, RSJ Frackowiak, and RJ Dolan. The trouble with cognitive subtraction. *Neuroimage*, 4(2):97–104, 1996.

[6] Ron D Frostig, Edmund E Lieke, Daniel Y Ts'o, and Amiram Grinvald. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by in vivo high-resolution optical imaging of intrinsic signals. *Proceedings of the National Academy of Sciences*, 87(16):6082–6086, 1990.

[7] Gary H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9(4):416 – 429, 1999.

[8] Daniel A Handwerker, John M Ollinger, and Mark D'Esposito. Variation of bold hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–1651, 2004.

[9] M Lotze, M Erb, H Flor, E Huelsmann, B Godde, and W Grodd. fMRI evaluation of somatotopic representation in human primary motor cortex. *Neuroimage*, 11(5):473–481, 2000.

[10] M-Marsel Mesulam. From sensation to cognition. *Brain*, 121(6):1013–1052, 1998.

[11] Thomas Nichols, Matthew Brett, Jesper Andersson, Tor Wager, and Jean-Baptiste Poline. Valid conjunction inference with the minimum statistic. *Neuroimage*, 25(3):653–660, 2005.

[12] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24): 9868–9872, 1990.

[13] S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle, and K. Ugurbil. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A*, 89(13):5951–5955, July 1992. ISSN 0027-8424.

[14] Seiji Ogawa, Tso-Ming Lee, Asha S. Nayak, and Paul Glynn. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, 14(1):68–78, 1990.

[15] Russell A Poldrack. 13 subtraction and beyond: The logic of experimental designs for neuroimaging. *edited by Stephen José Hanson and Martin Bunzl*, page 147, 2010.

[16] Cathy J Price and Karl J Friston. Cognitive conjunction: a new approach to brain activation experiments. *Neuroimage*, 5(4):261–270, 1997.

[17] Burrhus F Skinner. Operant behavior. *American Psychologist*, 18(8):503, 1963.

[18] William R Uttal. Response to bechtel and lloyd. *Brain and Mind*, 3(2):261–273, 2002.

2 Tools for neuroimaging data modeling

We have seen in Chapter 1 how researchers use functional MRI as an experimental tool to inspect brain functions. This chapter describes the statistical methods commonly used to analyse fMRI data, and introduces the different meta-analysis approaches used by the community to analyse jointly multiple studies.

The first section of this chapter presents the statistical inference framework used by the majority of individual fMRI experiments. It introduces the modeling of the fMRI BOLD signal, and the fundamental elements of hypothesis testing. We show how these methods yield activation maps representative of the tested cognitive process.

The second section describes the basics of statistical learning for fMRI. We introduce supervised linear predictive models in the context of fMRI, and common techniques to validate the models. We also describe dimension reduction methods, that serves to mitigate the issues raised by the high-dimensionality of fMRI data. We present the *decoding* approach, which is an application of predictive models that uses functional brain images to predict the engaged cognitive process.

The third and last section summarizes the meta-analyses methods for fMRI. We describe in particular coordinate-based meta-analyses, and image-based meta-analyses.

Contents

2.1 Statistical inference

This section describes the standard framework to perform statistical inference on BOLD neuroimaging data. BOLD images typically go through a number of pre-processing steps, which include different corrections, smoothing, and normalization to a common brain space. We will not describe these procedures here, but rather describe the statistics underlying BOLD modeling. More details on pre-processing procedures may be found in Poldrack et al. [29].

2.1.1 Modeling fMRI data

Friston et al. [9] introduced the *general linear model* (GLM) to the neuroimaging community, which is the cornerstone of fMRI analysis. This approach yields statistical parametric maps that are used to test for region specific effects relative to a given hypothesis. This method can take into account all the different factors that explain the fMRI timeseries: the experimental paradigm, the physiological effects, some confounding variables, and noise. The result is *activation maps*, that assign effect estimates at each voxels for the different factors.

General Linear Model Let $\boldsymbol{X} \in \mathbb{R}^{n \times k}$ be the design matrix in which we represent the different factors as columnwise stacked regressors, where n is the number of scans, and k the number of regressors. The regressors consist of the experimental conditions (occurence of stimuli in the experimental design convoluted with the *HRF*), nuisance confounds such as subject motion, additional coufounds such as session or study-dependent effects, as well as low-frequency signals that model drifts in the signal. The time derivative of the experimental conditions is also often included, to offer some flexibility to the *HRF* modeling, which differs in delay and shape across the brain.

We denote as $Y \in \mathbb{R}^{n \times p}$ the observed fMRI data, where n is the still the number of scans, and p the number of voxels. The GLM formulation gives us:

$$Y = X\beta + \varepsilon \quad , \tag{2.1}$$

where $\beta \in \mathbb{R}^{k \times p}$ are the effects, and ε the residual error. The error term ε is assumed to be drawn from a normal distribution $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. Figure 2.1 represents the different terms in the GLM formulation. Importantly, the effects term may be represented as brain images, and are called *activation maps* or $\hat{\beta}$ *maps*. Figure 2.2 shows the estimated $\hat{\beta}$ *maps* of an experiment considering visually presented calculation and sentences tasks.

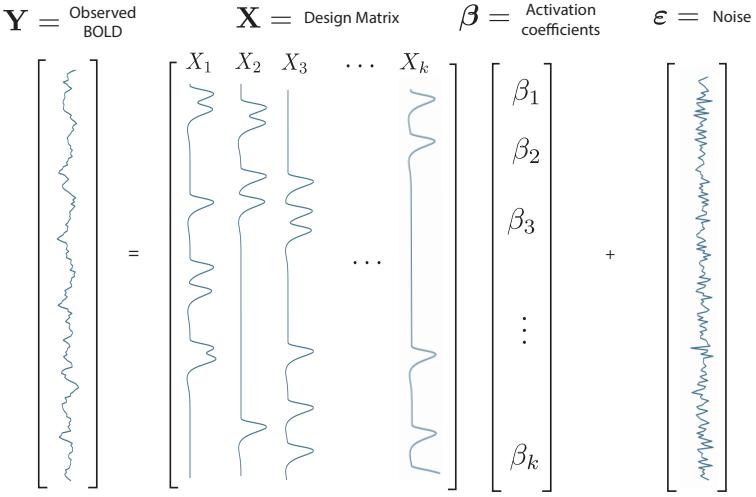


Figure 2.1: The GLM postulates that the observed data are a combination of effects to be estimated plus a noise. The regressors in the design matrix consist of the modeling of experimental conditions by the convolution of the stimuli presentation by the *HRF*, as well as confounding variables and filters. Source: Pedregosa-Izquierdo [24]

We introduced here the modeling of individual effects, *i.e.* the estimation of the activation maps at a subject-level from its BOLD images. The standard framework for fMRI data analysis usually follows a hierarchical structure, in which researchers model the effects at a *group-level* after computing the maps at the *subject-level*. Group-level maps are estimated with the same GLM framework, but take as inputs the subject-level maps instead of the BOLD timeseries.

Hypothesis testing In this paragraph, we present the general setting of hypothesis testing and it is instantiated in the particular context of fMRI. Statistical hypothesis testing is the process of testing a hypothesis on the distribution of a given parameter. In cognitive neuroscience, researchers designing experiments test whether the observed brain activity is due to chance. If not, the test concludes the activity is related to the experimental conditions. In the general setting of statistical inference, we call *null hypothesis* an observation occurring by chance. The null hypothesis is considered true until sufficient evidence imply otherwise, in which case it is rejected in favor of the *alternative hypothesis*. The *null hypothesis* and the *alternative hypothesis* are respectively denoted H_0 and H_1. The *p-value* resulting from a hypothesis testing corresponds to the probability of an observation

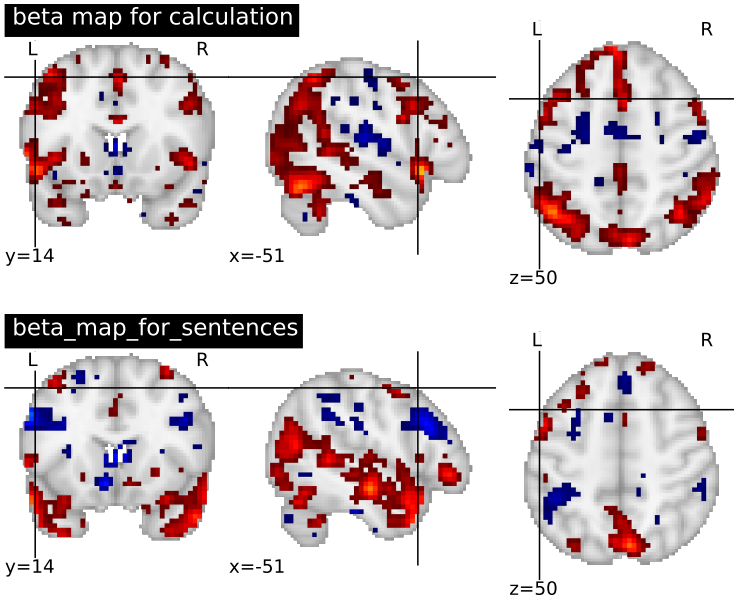


Figure 2.2: Beta maps for visually presented calculation and sentences tasks. The thresholding is arbitrary and for visualization purpose only. Source: Pinel et al. [25]

.

considering H_0 is true (see Figure 2.3). It enables to define the *significance level*, which is an arbitrary threshold at which we disprove $H0$. This threshold is traditionally set at 5% or 1%, and guarantees that the type I error is no greater than the chosen threshold. Type I and type II errors are also respectively referred to as false positives and false negatives.

A t-test is a statistical test in which the test statistic follows a Student's t distribution under the null hypothesis [32]. In the following, we consider the one-sample t-test, whose statistic with a sample $\boldsymbol{x} \in \mathbb{R}^n$ is given by:

$$t = \frac{\bar{\boldsymbol{x}} - \mu_0}{s/\sqrt{n}} \quad , \tag{2.2}$$

where $\bar{\boldsymbol{x}}$ is the sample mean, s the sample standard deviation, and n the sample size. A one-sample t-test has the null hypothesis that the population mean is equal to the specified value μ_0. With μ the population mean, a one-sample t-test hypothesis may therefore be formulated as:

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0$$

Once the t value is determined, the corresponding p-value may be found using the Student's probability distribution with $d = n - 1$ degrees of freedom. P-values may be computed under a *one-tailed* or a *two-tailed* test, depending on whether the deviation of the estimated parameter may be on either direction of the sample mean, or only one direction. In fMRI we generally use a *two-tailed* test, as we are interested in significant activations, as well as deactivations. Figure 2.3 shows a probability distribution with the one-tailed test highlighted in green. A two-tailed test would also consider a symmetrical area on the other side of the distribution.

Statistical parametric maps Cognitive neuroscientists aim to isolate mental processes of interest. They do so by contrasting a stimulus engaging the mental process, with a stimulus theoritically identical in all aspects but the

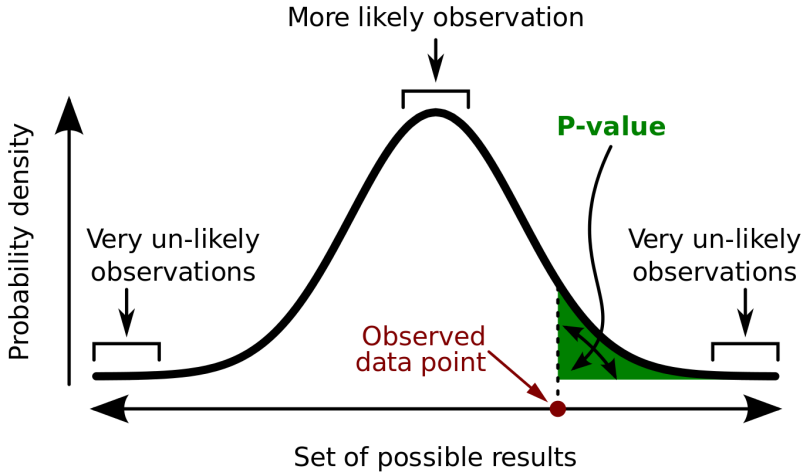


Figure 2.3: A p-value (green area) is the probability of an observed result assuming H_0 is true under a one-tailed test. Source: modified from `http://en.wikipedia.org/wiki/File:P-value_in_statistical_significance_testing.svg`

aforementioned process. For example, studying *calculation* implies presenting an arithmetic operation to a subject, *e.g.,* "two plus six". If this stimulus is visual, it will recruit regions from the visual system, as well as from the reading process. To identify, brain activations specific to calculation, we use a sentence of the same length as a control condition to cancel out the undesired effects, *e.g.,* "less is more". The difference between a condition and its control serves as the null hypothesis to find significant activations relative to calculation. We introduce the notion of a *contrast* $c \in \mathbb{R}^k$ as a linear combination of experimental conditions exhibiting a mental process. Given the GLM equation 2.1, we define H_0 as $\boldsymbol{c}^T \boldsymbol{\beta} = 0$ and H_1 as $\boldsymbol{c}^T \boldsymbol{\beta} \neq 0$. This tests whether the $\boldsymbol{\beta}$ of the condition and its control are significantly different, and the t-statistic is given by:

$$t = \frac{\mathbf{c}^T \hat{\boldsymbol{\beta}}}{\hat{\sigma} \sqrt{\boldsymbol{c}^T \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{c}}} \quad , \tag{2.3}$$

where $\hat{\sigma}^2$ is the estimate of the variance, and the Student's distibution has $d = n - (k + 1)$ degrees of freedom. Because there is one t-test per voxel, we obtain a t-map representing the brain activity with one t-statistic per voxel. As an example, we consider an experiment with two conditions: calculation and sentences. The GLM gives the $\hat{\boldsymbol{\beta}}$ *maps* from Figure 2.2, with a design matrix containing two columns for the experimental conditions, and the additional columns for confounding variables. In this case the contrast vector c takes the form $\boldsymbol{c} = [+1, -1, 0, \ldots, 0]$, where $+1$ is for the calculation regressor, and -1 for the sentences one. The resulting t-map in Figure 2.4 is thresholded at $t = 3.2 \sim pvalue = 1‰$, and shows activations in the parietal cortex relative to calculation. It also cancels out activations in the visual and language systems that can be seen in the $\hat{\boldsymbol{\beta}}$ *maps* from Figure 2.2. These kind of maps are more generally called statistical parametric maps (SPMs). At the *group-level* it is possible to compute these *contrast maps* with a two-sample t-test.

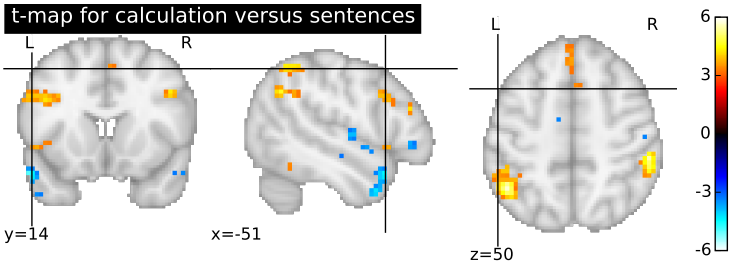


Figure 2.4: t-map for calculation versus sentences thresholded at $t = 3.2$. This contrast map cancels out effects related to the stimulation rather than the calculation task itself. Source: Pinel et al. [25]

.

2.1.2 Multiple comparisons problem

The multiple comparisons problem arises when one considers a set of statistical tests simultaneously. Let α be the significance level, with $\alpha = 5\%$ a single test ensures that the probability to reject incorrectly the null-hypothesis is 5%. This controls the type I error. Let m be the number of tests, and s the number of significant results, assuming independence of the tests the probability of observing at least one significant result due to chance is given by:

$$\mathcal{P}_{s\geq 1} = 1 - \mathcal{P}_{s=0}$$

$$\mathcal{P}_{s\geq 1} = 1 - (1-\alpha)^m$$

which means the probability to observe at least one false positive with 20 tests is 64%, and 100 tests is 99.4%. An fMRI image with voxels at a $3mm^3$ resolution contains around 50K voxels and corresponds to the number of tests we perform. Since we want to answer a *where* question, *i.e.* localize activations for a given function, it is critical to control correctly the type I error rate. The proposed solution is to control for an experiment-wide significance level instead of a single test significance level, and is called family-wise error rate (FWER). The general idea to obtain a specified FWER, is to define error rates for each comparison that are more stringent than α. There are several procedures to account for the multiple comparison problem. One of them is the Bonferroni correction, which sets the new significance level to $\alpha_b = \alpha/m$, and implies that the global error rate will not be greated than α. The Bonferroni method may be too conservative in fMRI analysis, as is do not take into account the high correlation of neighboring voxels. Other correction methods that may be more adapted to fMRI include random field theory [40, 8], and the use of the false discovery rate [1].

2.2 Statistical learning for fMRI

Machine learning is a scientific field that encompasses a set of algorithms that learn from data. The algorithms operate on inputs and outputs to build models, which are used to make decisions. There are several classes of algorithms, such as supervised and unsupervised learning, which depend on the type of task to be performed. Machine learning has originally been used in neuroimaging to uncover the neural coding relative to a task, *i.e.*, the voxels predictive of a mental process. This pratice is referred to as *decoding* or multivoxel pattern analysis (MVPA) in the neuroimaging community

[5, 4, 15, 34, 31]. The main incentive for this kind of approach is that is uses a single statistical test for the whole brain, and thus alleviates limitations due to multiple comparisons in classical inference. More recently, Poldrack et al. [28] has shown that decoding enables to implement a large scale *reverse inference*, in order map cognitive functions to brain areas.

2.2.1 Predictive models

Supervised predictive modeling is a set of statistical learning methods which aim to predict a target variable. In fMRI, we consider the activation maps produced by the GLM as a pattern recognition problem, *i.e.* predict a cognitive state, also called *target*, using a pattern of voxels. Note that is it also possible to perform such predictions using voxels from the fMRI timeseries, but we do not perform this kind of analysis in this thesis. Predictive models learn a *predictive function* which which can be linear or non-linear. Non-linear functions can in principle uncover non-linear relationships between the voxels and the target, but their predictive power has not been found to be clearly exceeding that of linear models [13] for brain analysis. More importantly, our goal is to map the predictive function onto the brain, which is trivial with a linear function. This is also feasible with non-linear functions [21, 22, 20, 17] but is less frequently used, which is why we focus only on linear models in this thesis like the majority of the neuroimaging community [4, 15, 34, 31]. The advantages of a predictive model over the classifical statistical inference framework presented in the previous section are that: i) it performs a single statistical test for all the voxels, and ii) its multivariate nature enables it to capture distributed patterns that are predictive of a given target. The latter point is crucial, as brain regions may explain a mental process if considered in conjunction, but may overlap if taken separately.

Linear models Linear predictive models can be used for classification and regression problems. Classification predicts a discrete target composed of *classes*, whereas regression predicts a continuous variable, *e.g.*, age. We only consider classification in this thesis. Let $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ be the data, with n the number of samples and p the number of voxels in each activation map, and $\boldsymbol{y} \in [1, \dots, k]^n$ the target composed of k classes. With these notations, a linear predictive model for classification is given by:

$$\boldsymbol{y} = f(\boldsymbol{X}, \boldsymbol{w}, b) = F(\boldsymbol{X}\boldsymbol{w} + b) \quad , \tag{2.4}$$

where $\boldsymbol{w}$ and b respectively are the *slope* (weights) and the *intercept* of the affine hyperplane. The intercept serves to choose an offset that is different from 0 for the affine hyperplane. Following this formulation, we define a binary classification as:

$$f(\boldsymbol{X}, \boldsymbol{w}, b) = \text{sign}(\boldsymbol{X}\boldsymbol{w} + b) \quad , \tag{2.5}$$

where sign is the *sign* function and $y \in [-1, 1]^n$. The estimation of the parameters is done through the minimization of the difference between y and the estimated target $\hat{\boldsymbol{y}}$. This difference is a function of $(\boldsymbol{w}, b)$ called *loss function* –or *loss*– that we denote $\ell(\boldsymbol{w}, b)$. The loss represents the cost

associated with the error in the estimation of the target. In classification settings the most common losses are the *hinge* and *logistic* losses. We mostly use the logistic loss in this thesis, that we express as follows:

$$\ell(\boldsymbol{w}, b) = \frac{1}{n} \sum_{i=1}^{n} \log\left(1 + \exp^{-y_i({X_i}^T \boldsymbol{w} + b)}\right) \tag{2.6}$$

This type of loss is used by the *logistic regression* model, which models the posterior probability as a sigmoid: $\mathcal{P}(y|X) = \left(1 + \exp^{-yf(x)}\right)^{-1}$. It enables to associate class probabilities with the data and ultimately assign classes. Other models such as Support Vector Machines (SVM) use different loss functions but follow the same principles.

Regularization *Regularization* uses prior knowledge to bias a model's estimation towards a desired solution. It enables to solve ill-posed problems and avoid overfitting. With $\lambda\,\Omega(\boldsymbol{w})$ the regularization term, the minimization problem becomes:

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w},b}{\arg\min} \;\; \ell(\boldsymbol{w}, b) + \lambda\,\Omega(\boldsymbol{w}) \;\;, \quad \lambda \geq 0 \tag{2.7}$$

where λ represents the amount of penalization, and is the tradeoff between the loss $\ell(\boldsymbol{w}, b)$ and the penalty $\Omega(\boldsymbol{w})$. Among the many possible penalties, we present the most widespread ones in neuroimaging:

- ℓ_2 regularization ($\Omega(\boldsymbol{w}) = \|\boldsymbol{w}\|_2$). $\|\boldsymbol{w}\|_2$ is the Euclidian norm, and this type of regularization is equivalent to setting a Gaussian prior with zero mean on $\boldsymbol{w}$.
- ℓ_1 regularization ($\Omega(\boldsymbol{w}) = \|\boldsymbol{w}\|_1$). $\|\boldsymbol{w}\|_1 = \sum_{i=1}^{n} |x_i|$ is the Manhattan norm, and this type of regularization forces a large fraction of uninformative weights to zero, promoting *sparse* models.

Model validation and selection A model's validation is done throught the evaluation of its predictive performance. To avoid overfitting the standard procedure is to partition iteratively the data into a *training* and a *testing* set, where we learn the prediction function on the training set and evaluate it on the testing set. As the prediction scores may vary across training and testing sets, we usually report the averaged prediction scores (and their standard deviation). This procedure is called *cross-validation*. The main pitfall of cross-validation is that it holds out part of the data for validation, which reduces the number of samples for the learning of the model. For most use cases of machine learning this is not an issue, but may be so in the case of a single neuroimaging study which features few samples. We here describe a few popular cross-validation schemes, where the schemes names are taken from *scikit-learn*[1] and may not be representative of all the literature:

- *Leave-one-out*: the training set is composed of all samples but one held out for testing. Given n samples, this cross-validation scheme therefore generates n different training and testing splits.

[1] `http://scikit-learn.org/stable/modules/cross_validation.html`

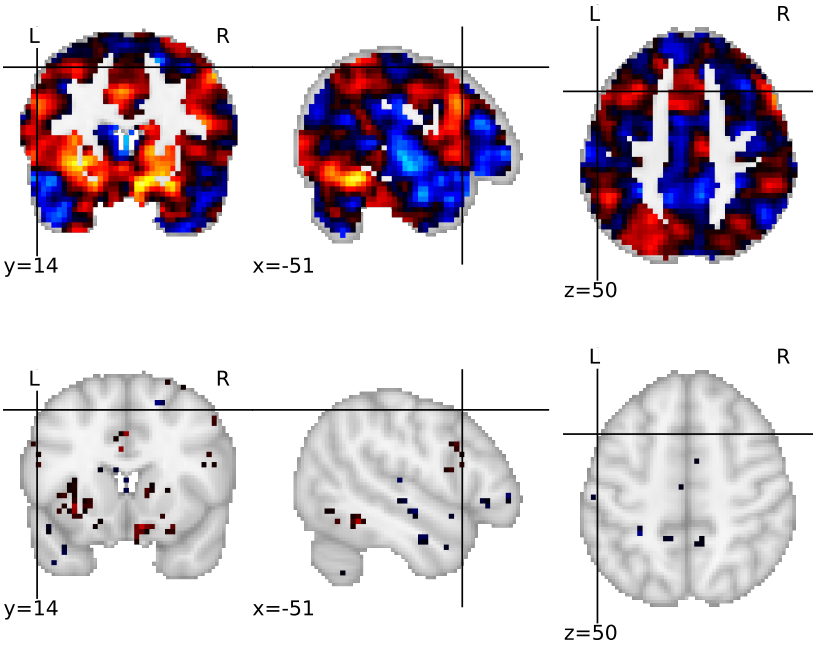


Figure 2.5: This figure shows the coefficients of a logistic regression discriminating between a calculation and a sentence reading task. The logistic regression is penalized with an ℓ_2 regularization term on the top row, and an ℓ_1 regularization on the bottom row. We choose the amount of penalization by cross validation for both models and kept the best performing model. The ℓ_1 penalty yields a very *sparse* model, *i.e.* relies on very few predictive features.

- *K-fold*: this scheme partitions the data into K disjoints group of the closest size called *folds*. The training set is composed of $K-1$ folds and the testing set uses the remaining one.

- *Shuffle-split*: this scheme returns random but disjoint training and testing sets. The training and testing sizes, as well as the number of desired splits, are user defined parameters of the cross-validation, e.g. 5 random testing sets representing 10% of the data.

- *Stratified-shuffle-split*: this scheme is identical to *shuffle-split* but ensures that the distribution of classes is the same in the training and testing sets. This is useful in the case of *class imbalance*, to avoid biasing the estimator in favor of the majority class. We implemented a version of stratified-shuffle-split for *scikit-learn* for the purpose of this thesis.

- *Leave-one-label-out*: this scheme is similar to the *leave-one-out* but holds out the samples based on third-party labels, which define a domain specific cross validation. For example, in neuroimaging it is common to perform a leave-one-subject-out cross validation to test whether a given prediction function generalizes across subjects, and does not rely on subjects' idiosyncrasies for prediction. A leave-one-label-out may be used to test for all possible confounding effects that we want to test for, *e.g.*, session-effect, study-effect, scanner-effect, or site-effect.

We may use different metrics to estimate the performance of a model within a cross-validation loop. For a classification task, those metrics use the predictions of the classifier which may be true positives (TP), true negatives (TN), false positives (FP), or false negatives (FN). Follows a description of the main metrics we consider in neuroimaging, with P and N being simply positives and negatives, regardless of the correctness of prediction.

- *Accuracy*: $ACC = \frac{TP+TN}{P+N}$
- *Precision*: $PRC = \frac{TP}{TP+FP}$
- *Recall or sensitivity*: $REC = \frac{TP}{TP+FN}$
- *F1-score*: $F1 = \frac{2TP}{2TP+FP+FN}$

Accuracy is typically used in the case of a balanced classification and represents the rate of correct predictions. In the case of an imbalanced problem, accuracy may be misleading as it can yield a very high score by ignoring the minority class. To avoid this, it is preferable to consider for each class separately both precision and recall scores, which respectively represent the ability of a classifier to make correct predictions, and its ability to detect all the instances of one class. The F1-score is the harmonic mean of precision and recall and serves to have a quick overview of a classifier's performance.

Models usually depend on internal parameters to fit to a problem, *e.g.*, the amount of penalization. The process to select the best model, *i.e.*, the best set of parameters, for a given problem is called *model selection*. We select the model in respect to its prediction performance, and do so in a *nested* cross-validation loop to avoid overfitting. This procedure may give rise to new challenges, as it further partitions the data to generate (inner) training and testing sets, and is computationally expensive.

Multi-class and multi-label classification A *multi-class* classification is the problem of classifying data within more than two classes. By contrast, a *multi-label* classification assigns a set of target labels to each instance of the data. Multi-class problems resort to **voting strategies** to combine multiple binary classifiers. Given K classes, common approaches are:

- *One-versus-all*: also called one-versus-rest, this approach builds K binary classifiers by pooling the data from $K-1$ classes. Given $f(\boldsymbol{X}, w, b)$ from equation 2.4, the voting stategy consists in taking the highest real-valued confidence score among the binary classifiers. If $g(x)$ is the decision function and $\text{sign}(g(x))$ the binary prediction function, the multi-class prediction function is:

$$\hat{\boldsymbol{y}} = \underset{k \in 1 \dots K}{\arg\max}\, g_k(\boldsymbol{x})$$

  The advantage of this method is that in addition of being computationally efficient, it yields one decision boundary per class, which makes it easy to interpret.

- *One-versus-one*: this method considers all the possible pairs of binary classifiers, *i.e.* $K(K-1)/2$ classifiers. The voting strategy consists in predicting the class that has the majority of predictions among all the binary classifiers. In comparison to one-versus-all, it quickly becomes computationally expensive as K gets larger. Moreover, there is no straightforward way to retrieve one decision boundary per class.

For multi-label problems only the *one-versus-all* strategy applies. In this case we do not apply a voting strategy, but rather assign multiple classes to each instance of the data.

2.2.2 Dimension reduction

Learning a decision function requires more data samples as the dimensionality of the feature space gets higher. With a fixed number of samples, increasing the number of dimensions of the data may increase in turn the prediction accuracy up to an optimal point, after which it will reduce the accuracy [13]. This effect is known as the *Hughes effect*, or the *curse of dimensionality*. The underlying problem of a high-dimensional space, is that the volume of this space makes the data sparse. In other words, we may find an infinity of separating hyperplanes as solutions of a predictive model. Most of these solutions will however not be able to generalize to new samples (*e.g.*, in the testing set of a cross-validation), since it relied on specificities or noise from the training set. This problem is particularly severe in neuromaging, as typical studies feature around $n = 100$ activation maps of approximately $p = 50K$ voxels. Dimension reduction methods overcome these issues by defining a low-dimensional space keeping the predictive information. Dimension reduction is routinely used in neuroimaging, and is performed within the cross-validation loop to avoid any overfit. The most frequent approach is *feature selection*, which removes non-relevant features from the feature space. Another approach is *feature agglomeration*, which combines features to create a lower-dimensional space. We present here one *univariate* feature selection method, and one feature agglomeration method. In this thesis we usually combine both approaches to create a new feature space on which we perform the classification. Figure 2.6 illustrates this procedure.

Feature selection A common way to select features is to apply a so-called *univariate screening* on the data. We compute a score value for each feature independently, *e.g.* F-score, correlation score. We perform the selection itself by either thresholding the scores, keeping the k best scores, keeping a percentile of the highest scores, or by testing for significance with a p-value. The computational efficiency is the major advantage of this approach. It however does not take into account the local correlation of features, in our case the voxels which share information with their neighbors. Because of the thresholding, clusters of voxels with a high signal to noise ratio (SNR) are likely to be selected, while voxels that may still be informative at the classification level but have a lower SNR are likely to be ignored. This is why we use this selection method in a very non-conservative way, and usually keep at least 30% of the features.

Feature agglomeration The main limitation of a univariate feature selection for neuroimaging data, is that it does not take into account any spatial information. Feature agglomeration replaces the voxel-space by a parcel-space which summarizes the grouped voxels. This approach reduces the dimensionality and increases the SNR at the expense of spatial resolution, but does not exclude potentially informative features. Different criteria may be used to define the parcels. Using anatomical atlases may be the simplest way to do so, but are ill-suited to represent functional data because it yields a very coarse representation of the brain activity ($\sim$ 100

regions). We instead consider the Ward clustering algorithm [19], which is a clustering algorithm that uses the activation maps to learn a partition of the brain. Because it adapts the size of the parcels to the signal in the brain regions (large regions with similar values yield a single large parcel), the Ward clustering should be seen as a *lossy compression* algorithm, rather than a method yielding plausible functional regions. Our experience show that keeping between 2K and 5K parcels does not hinder classification, and makes computation faster.

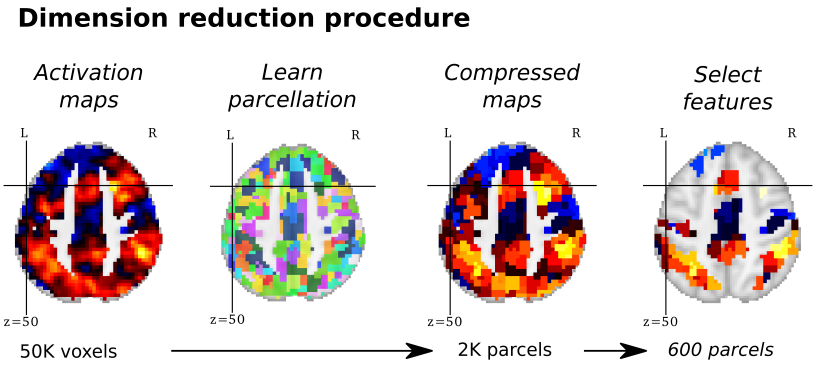


Figure 2.6: The Ward feature agglomeration algorithm learns a partition of the brain composed of 2K parcels, on which we sample the activation maps. We obtain *compressed* activation maps and proceed with a univariate screening which keeps 30% of the features to further reduce the dimensionality of the data.

2.2.3 Decoding

Predicting mental processes Decoding in neuroimaging is the process of predicting mental states using a supervised learning model. Early proofs of concept show the ability of a classifier to accurately predict left from right hand movements [20, 5], and different categories of visual stimuli (faces, houses, objects) [11, 4], as well as uncovering the underlying neural coding of these processes. Later work demonstrates that decoding may also be used to expose shared neural support as in *neuronal recycling*, by training a classifier on one task and predicting another one which shares the same neural basis with the same classifier [14]. The validity of decoding have also been shown on higher-level mental concepts such as working memory [10] or intentions [12].

Stability selection Cross validation techniques validates the prediction accuracy but neither the validity, nor the interpretability of the classifier's coefficients. Decoding approaches are usually regarded as more powerful than their univariate counterparts because they consider all voxels at the same time. They however do not perform the same statistical test, and do not control for any error rate at the voxel level. As a consequence, a classifier may assign high weights to irrelevant features. This issue is related to model selection and the curse of dimensionality previously described. The process of finding stable features that are representative of a task is known as *stability selection* [18]. In this thesis, we rely on an approach akin to the one developed in Varoquaux et al. [38]. Specifically, we generate randomized parcellations using a Ward clustering, and select the best model varying the regularization parameter for each those parcellations. Each model therefore has a coefficient vector in the parcel-space. We project back the best models to the voxels-space and average them to obtain a final linear

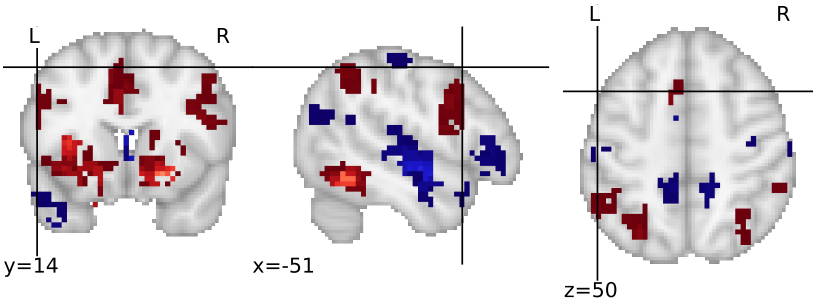


Figure 2.7: This hyperplane map shows the coefficients of an averaging model, which are more interpretable than simple ℓ_2 or ℓ_1-penalized models (see Figure 2.5). In particular the high coefficients in the parietal cortex are relevant for a calculation task.

model in the voxel space. As in Varoquaux et al. [38], we choose an ℓ_1 penalty for the linear classifiers. This method proved to be easy to implement, relatively efficient in terms of computation, and yields more stable and interpretable coefficients. Using this method for the same calculation versus sentences classification, we obtain the coefficients pictured in Figure 2.7, that are more specific than a simple ℓ_2-penalized logistic regression, and less sparse than an ℓ_1-penalized model.

2.3 Quantitative meta-analyses

Meta-analyses comprise statistical methods for combining and contrasting different studies. The original motivation behind meta-analyses is to aggregate information from multiple studies to increase the statistical power for a measure of interest. It is also a way to integrate and summarize results on a specific topic. The functional MRI field faces a vast and rapidly growing literature on various cognitive domains, and can benefit from approaches that join seemingly unrelated topics. Meta-analyses of fMRI have additionally the potential to overcome some limitations of individual studies, in particular false positive and negative findings. More generally, meta-analyses are an interesting tool to develop new hypotheses, assess the consistency across experimental protocols, or reach consensus on the location of functional regions. This section presents two kinds of meta-analyses for fMRI: the coordinate-based meta-analyses which use functional activation summaries, and image-based meta-analyses that use functional brain images.

2.3.1 Coordinate-based meta-analyses

Coordinate-based meta-analysis (CBMA) methods assess the convergence of activation peak coordinates across multiple experiments to synthetize, reconcile, or develop findings from the literature. They would not be feasible without the high standardization of neuroimaging reports. Most neuroimaging data is standardized to either the Talairach-Tournoux [33] or Montreal Neurological Institute (MNI) [7] brain spaces. These common coordinate systems make it possible to compare of brains across subjects and studies. Results of neuroimaging studies report tables of peak coordinates for all activations of interest. Peaks are defined as the local maximum of an activation given by the comparison of two experimental conditions. This makes potentially all the literatue eligible to be part of such meta-analyses, as papers may be entered manually in a coordinate database, or tools may

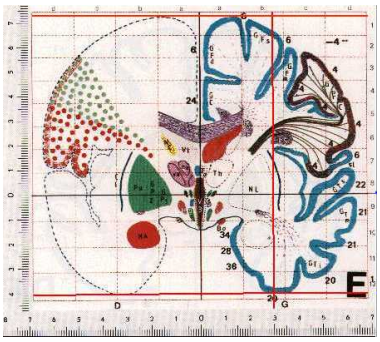


Figure 2.8: The Talairach atlas is the first instance of a coordinate system for the brain. It was created by neurosurgeons Jean Talairach and Gabor Szikla from the post-mortem dissection of a single human brain. Source: `http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach`

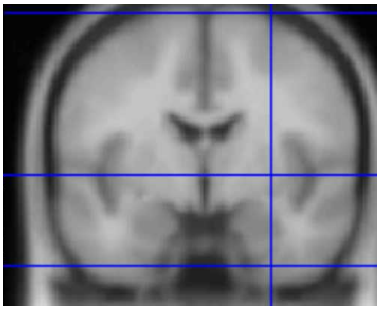


Figure 2.9: The current MNI template is the ICBM152. It comes from the average of 152 MRI scans matched to the MNI305, which is the first MNI template. This first template was obtained by averaging and matching 305 MRI scans to structures from the Talairaich atlas. Source: `http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach`

be developed to integrate automatically vast fractions of the literature. A standard format for table reports [27] would facilitate the latter.

A taste of ALE Activation Likelihood Estimation is a popular method for CBMA introduced by Turkeltaub *et al.* [36], and later refined by Eickhoff *et al.* [6]. The gist of this method is to treat each activation peak as the center of a Gaussian probability distribution, to account for the spatial uncertainty entailed by the between subject and laboratory variances. A 3D modeled activation map (MA map) summarizes each experiment in a common brain space, and enables to create a final ALE map which represents the union of all MA maps. The final step of this method is to assess the statistical significance of each location indicated by the ALE map. The original method [36] proposes to test for above-chance clustering of peaks, whereas the current method [6] tests instead for the above-chance clustering of experiments, *i.e.* MA maps. ALE therefore defines the null hypothesis as the random association of peaks between MAP maps, and the alternative hypothesis as the convergence of peaks for a given task. This testing enables to draw conclusions on a mental process of interest, as shows Figure 2.10 for finger tapping. This method is also used to find functional modules in brain structures, to achieve a finer representation of functions in the brain [2].

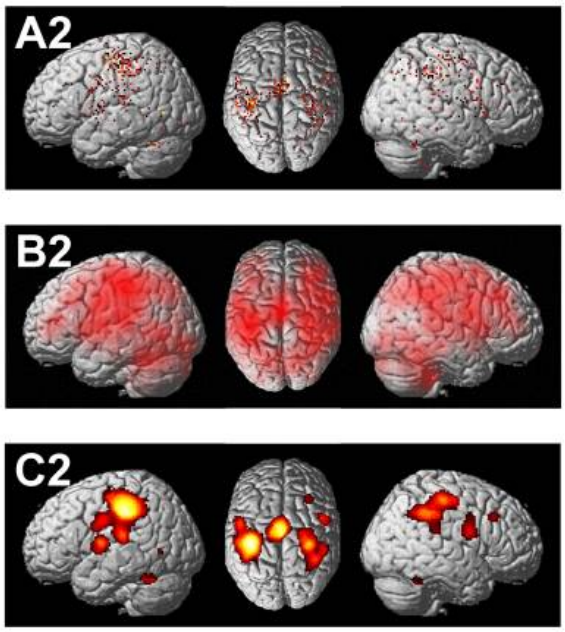


Figure 2.10: A2: individual activation peaks for the finger tapping experiments included in the meta-analysis. B2: Union of the MA maps across all experiments. C2: significant cluster across experiments against a null distribution obtained by permutation testing. Source: Eickhoff et al. [6]

The ALE method is applicable in any CBMA, but is primarily associated with the BrainMap database [16]. BrainMap features brain locations from more than 2,500 papers encompassing over 12,000 experiments. All papers and annotations are manually entered in the database, which makes it the largest of this kind. BrainMap also exposes a collection of online tools, in particular GingerALE [6] which enables to perform ALE analyses with data from the database, or from a manually uploaded table of coordinates. Other "manual" coordinate databases include SumsDB[2] [37] and Brede Wiki[3] [23].

[2] `http://sumsdb.wustl.edu:8081/sums/index.jsp`

[3] `http://neuro.compute.dtu.dk/wiki/Main_Page`

Neurosynth BrainMap focuses on the quality and accuracy of the papers annotations entered in the database. Despite the large representation of the literature in its database, it still represents a fraction of all potential studies, and depend on inclusion criteria that are not explicit. Neurosynth[4] [41] takes the other approach, and pools automatically as many studies as possible from the literature in order to have the best representation, and relies on the volume of data to make up for poorer annotation quality. In its original 2011 version, Neurosynth was drawing activation foci from almost 3, 500 papers, its current version is now over 10, 000. Neurosynth automated coordinate extraction consists of a parsing engine that detects tables reporting activation coordinates, and performs basic validation. The differences in standard spaces for coordinates are ignored by Neurosynth. Moreover, Neurosynth parses the article text and performs a word frequency analysis to generate lists of terms to associate with the coordinates. This fully automated approach enables to construct a large database of term to coordinates mapping, and is able to cope with the ever growing load of publications. Neurosynth also provides forward and reverse statistical inference tools for the analysis of its database. The forward inference method tests for the dependence between terms and activation using a χ^2 test. Yarkoni et al. [41] justifies the use of a parametric statistical test rather than permutation testing mainly for computational purposes. Reverse inference is implemented as a naive Bayes classifier. By computing the probability of a term given an activation, this classifier is able to predict terms on statistical maps from unseen neuroimaging studies. Reverse inference on such a database is of particular interest, as it covers extensively the cognitive space.

[4] http://neurosynth.org/

Limitations CBMA aim to synthesize large amounts of neuroimaging experiments, and overcome limitations inherent to individual studies, in particular false positive (FP) and false negative (FN) results. Most studies rely on voxel-wise hypothesis testing by rejecting the null hypothesis. The FP rate is in general arbitrarily set to 5% and corrected for multiple comparisons with the Bonferroni correction for example. FP may still be reported in the literature in the form of uncorrected results, and it is an established practice to perform region of interest analysis to reduce the severity of corrections [26]. CBMA mitigate FP results as they are hopefully not replicated across studies, even though being present in 10 to 20% of publications [39]. They however combine results obtained with potentially very different methods, and do not provide a way to account for these differences [17]. Reports show that individual studies are typically underpowered [35], and thus fail to detect engaged regions, otherwise known as FN results. This issue proves more problematic for CBMA methods, which represent non-significant study-level voxels as zeros, and therefore cannot aggregate power across studies to potentially reach significance [3]. The last major limitation from CBMA approaches is that of spatial resolution: activation peaks are modeled roughly as balls, and do not take into account the shape and size of activations, which are likely to greatly vary depending on the mental process under study.

2.3.2 Image-based meta-analyses

Pooling fMRI images have the potential to overcome several CBMA shortcomings. [3, 30] distinguish two kind of analyses performed by aggregating images, image-based meta-analyses (IBMA) and "mega-analyses". IBMA are performed on the full statistical images, resulting either from a subject-level analysis or a group-level analysis. The use of these data can give greater details on the shape and size of the activations, and also enables to aggregate power across studies, as they contain the statistical values for all voxels, not only those that are significant. They also do not suffer from the FP reported in the literature [39]. They however do not alleviate issues that may come from the differing processing methods used to obtain those maps. To solve this, the raw images –the BOLD timeseries– must be processed homogeneously in the context of a "mega-analysis". There is no standard method for *mega-analyses*, but they enable to model inter-study variability as a random effect, and introduce a third level in a GLM framework (after subject and group levels) [30]. The main limitation with these approaches remains the difficulty to gather data from many experiments.

2.4 Conclusion

We have seen in this chapter the statistical tools that we will use along this thesis. We have also presented the current state of meta-analysis methods for fMRI. Meta-analyses hold the promise to provide an automated way to accumulate knowledge and relate cognitive fields. Coordinate-based meta-analyses have so far been the most successful approach, mostly because of the difficulty to accumulate imaging data, and despite the greater potential of image-based meta-analyses. In the next chapters, we will show how we aggregate imaging data, and how we leverage the statistical tools we presented to get an integrated view of the human brain functions.

Bibliography

[1] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.

[2] Mareike Clos, Katrin Amunts, Angela R Laird, Peter T Fox, and Simon B Eickhoff. Tackling the multifunctional nature of broca's region meta-analytically: Co-activation-based parcellation of area 44. *Neuroimage*, 83:174–188, 2013.

[3] Sergi G Costafreda. Pooling fMRI data: meta-analysis, mega-analysis and multi-center studies. *Frontiers in neuroinformatics*, 3, 2009.

[4] David D Cox and Robert L Savoy. Functional magnetic resonance imaging (fMRI)"brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19(2):261–270, 2003.

[5] Stanislas Dehaene, Gurvan Le Clec'H, Laurent Cohen, Jean-Baptiste Poline, Pierre-François van de Moortele, and Denis Le Bihan. Inferring behavior from functional brain images. *Nature neuroscience*, 1(7):549–549, 1998.

[6] Simon B Eickhoff, Angela R Laird, Christian Grefkes, Ling E Wang, Karl Zilles, and Peter T Fox. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human brain mapping*, 30(9):2907–2926, 2009.

[7] AC Evans, DL Collins, and B Milner. An mri-based stereotactic atlas from 250 young normal subjects. In *Soc. neurosci. abstr*, volume 18, page 408, 1992.

[8] Karl J Friston, Keith J Worsley, RSJ Frackowiak, John C Mazziotta, and Alan C Evans. Assessing the significance of focal activations using their spatial extent. *Human brain mapping*, 1(3):210–220, 1994.

[9] Karl J Friston, Andrew P Holmes, Keith J Worsley, J-P Poline, Chris D Frith, and Richard SJ Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping*, 2(4), 1996.

[10] Stephenie A Harrison and Frank Tong. Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238):632–635, 2009.

[11] James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539): 2425–2430, 2001.

[12] John-Dylan Haynes, Katsuyuki Sakai, Geraint Rees, Sam Gilbert, Chris Frith, and Richard E Passingham. Reading hidden intentions in the human brain. *Current Biology*, 17(4):323–328, 2007.

[13] G Hughes. On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on*, 14(1):55–63, 1968.

[14] André Knops, Bertrand Thirion, Edward M Hubbard, Vincent Michel, and Stanislas Dehaene. Recruitment of an area involved in eye movements during mental arithmetic. *Science*, 324(5934):1583–1585, 2009.

[15] Stephen LaConte, Stephen Strother, Vladimir Cherkassky, Jon Anderson, and Xiaoping Hu. Support vector machines for temporal classification of block design fMRI data. *NeuroImage*, 26(2):317–329, 2005.

[16] Angela R Laird, Jack J Lancaster, and Peter T Fox. Brainmap. *Neuroinformatics*, 3(1):65–77, 2005.

[17] Nicholas Lange, Stephen C Strother, Jon R Anderson, Finn Å Nielsen, Andrew P Holmes, Thomas Kolenda, Robert Savoy, and Lars Kai Hansen. Plurality and resemblance in fMRI data analysis. *NeuroImage*, 10(3): 282–303, 1999.

[18] Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

[19] Vincent Michel, Alexandre Gramfort, Gaël Varoquaux, Evelyn Eger, Christine Keribin, and Bertrand Thirion. A supervised clustering approach for fMRI-based inference of brain states. *Pattern Recognition*, 45(6):2041–2049, 2012.

[20] Niels Mørch, Lars K Hansen, Stephen C Strother, Claus Svarer, David A Rottenberg, Benny Lautrup, Robert Savoy, and Olaf B Paulson. Nonlinear versus linear models in functional neuroimaging: Learning curves and generalization crossover. In *Information processing in medical imaging*, pages 259–270. Springer, 1997.

[21] Niels JS Mørch, Ulrik Kjems, Lars Kai Hansen, C Svarer, I Law, B Lautrup, S Strother, and K Rehm. Visualization of neural networks using saliency maps. In *1995 IEEE International Conference on Neural Networks*, pages 2085–2090, 1995.

[22] Finn Årup Nielsen. *Visualization and analysis of 3D functional brain images*. PhD thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark, 1996.

[23] Finn Årup Nielsen. Lost in localization: A solution with neuroinformatics 2.0? *Neuroimage*, 48(1):11–13, 2009.

[24] Fabian Pedregosa-Izquierdo. *Feature extraction and supervised learning on fMRI: from practice to theory*. Theses, Université Pierre et Marie Curie, February 2015. URL `https://tel.archives-ouvertes.fr/tel-01100921`.

[25] Philippe Pinel, Bertrand Thirion, Sébastien Meriaux, Antoinette Jobert, Julien Serres, Denis Le Bihan, Jean-Baptiste Poline, and Stanislas Dehaene. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC neuroscience*, 8(1):91, 2007.

[26] Russell A Poldrack. Region of interest analysis for fMRI. *Social cognitive and affective neuroscience*, 2(1):67–70, 2007.

[27] Russell A Poldrack, Paul C Fletcher, Richard N Henson, Keith J Worsley, Matthew Brett, and Thomas E Nichols. Guidelines for reporting an fMRI study. *Neuroimage*, 40(2):409–414, 2008.

[28] Russell A Poldrack, Yaroslav O Halchenko, and Stephen José Hanson. Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychological Science*, 20(11):1364–1372, 2009.

[29] Russell A Poldrack, Jeanette A Mumford, and Thomas E Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.

[30] Gholamreza Salimi-Khorshidi, Stephen M Smith, John R Keltner, Tor D Wager, and Thomas E Nichols. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *Neuroimage*, 45(3):810–823, 2009.

[31] Sutao Song, Zhichao Zhan, Zhiying Long, Jiacai Zhang, and Li Yao. Comparative study of svm methods combined with voxel selection for object category classification on fMRI data. *PloS one*, 6(2):e17191, 2011.

[32] Student. The probable error of a mean. *Biometrika*, pages 1–25, 1908.

[33] Jean Talairach and Pierre Tournoux. Co-planar stereotaxic atlas of the human brain. 3-dimensional proportional system: an approach to cerebral imaging. 1988.

[34] Bertrand Thirion, Edouard Duchesnay, Edward Hubbard, Jessica Dubois, Jean-Baptiste Poline, Denis Lebihan, and Stanislas Dehaene. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, 33(4):1104–1116, 2006.

[35] Bertrand Thirion, Philippe Pinel, Sébastien Mériaux, Alexis Roche, Stanislas Dehaene, and Jean-Baptiste Poline. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage*, 35(1): 105–120, 2007.

[36] Peter E Turkeltaub, Guinevere F Eden, Karen M Jones, and Thomas A Zeffiro. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage*, 16:765–780, 2002.

[37] David C Van Essen. Lost in localization—but found with foci?! *Neuroimage*, 48(1):14–17, 2009.

[38] Gael Varoquaux, Alexandre Gramfort, and Bertrand Thirion. Small-sample brain mapping: sparse recovery on spatially correlated designs with randomization and clustering. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 1375–1382. ACM, 2012.

[39] Tor D Wager, Martin Lindquist, and Lauren Kaplan. Meta-analysis of functional neuroimaging data: current and future directions. *Social cognitive and affective neuroscience*, 2(2):150–158, 2007.

[40] Keith J Worsley, Alan C Evans, S Marrett, P Neelin, et al. A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12:900–900, 1992.

[41] T Yarkoni, RA Poldrack, TE Nichols, DC Van Essen, and TD Wager. Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8:665, 2011.

Part II

Contributions: from an image database to learning brain functions

3 Scaling up from individual studies

The previous chapter focused on the analysis of individual fMRI studies, and the advent of large scale analyses using highly summarized data, i.e. the activation peaks coordinates. Dealing with individual studies makes it possible to test many models, and run quality assessment procedures at a subject-level if necessary. This approach is no longer tractable when the volume of data gets very large.

With large amounts of data comes new challenges, and this chaper investigates how to scale up from individual studies to a large fMRI database. We shortly review the available open-access fMRI resources, and how we contribute to that effort. We expose the strategy that we adopted to organize the data, and how it impacts the subsequent processing stages, in particular the fMRI pre-processing, statisical modeling, and data curation. We finish by giving an overview of the database that we accumulated.

Contents

3.1 Finding the data

The field of neuroimaging is seeing a shift in its datasharing policies, both thanks to a growing awareness of the opportunities it presents [36, 33], and the commoditization of the storage facilities. The interest in sharing fMRI data is not new, and was pioneered by the fMRI Data Center [46]. The platform closed in 2012 due to a lack of funding. Importantly, its major drawback when it operated was that it demanded an explicit request for each dataset before they are sent on a physical medium. The game changing initiatives are that of OpenfMRI [35], which features as of today as many as 29 task fMRI datasets, and the 1000 Functional Connectomes Project [4] which distributes over 1200 resting state datasets. Both projects make the data accessible by direct download under a permissive license. A major push to add new datasets is taking place for the OpenfMRI project, and many datasets were not available until recently which explains why we were not able to use all of them during the thesis. If large scale studies are not yet the norm in neuroimaging, they are becoming increasingly common and prove to be invaluable resources that set new datasharing standards for the community.The Human Connectome Project (HCP) [45] currently leads this trend by sharing the full data from over 500 subjects which can be either downloaded or ordered on a hard disk. Future projects include the European Human Brain Project (HBP), which holds a part dedicated to cognitive neuroscience and fMRI acquisition, and plans to acquire a vast range of experimental tasks on a very limited number of subjects. Past projects include the fBIRN [20], which was a US-wide multi-site project in the context of schizophrenia, and implemented datasharing on a large scale. Not all muti-site projects implement yet an open approach to datasharing, with the example of the IMAGEN project [38], that has the technical tools to share data from over 2000 subjects, but retricts access to a consortium.

Figure 3.1: The HCP shares data from over 500 subjects in 7 different fMRI tasks.

Figure 3.2: The OpenfMRI contains 29 datasets summing to 693 individual subjects.

Figure 3.3: Neurovault is an open web repository that eases the process to share statistical maps.

In a effort to reduce both sociological and technical frictions, we joined Krzysztof Gorgolewski's iniatiative to develop NeuroVault [16], a platform that aims to find the middle ground between sharing raw data, and only reporting the coordinates of the activation peaks in the papers. Neurovault.org is a web based repository that provides means to easily store, share, and visualize unthresholded statistical maps. The platform makes it trivial to upload a collection of maps, and link these with a permanent URL in the associated publication. The data are presently scarce, but growing quickly, and the ultimate ambition is to have enough data to be able to perform meta analyses, and foster new services that make use of its REST API [9].

Individual initiatives are also important and give access to high quality data. As part of the BioMag 2010 data competition, Henson released the raw fMRI, MEG, and EEG data of a faces recognition task [18].

In collaboration with the BrainOmics team from Neurospin and Logilab [1], we initiated a project to share neuroimaging data and associated metadata [25]. The result is a web repository containing data from [32, 31]. Specifically it contains the raw fMRI timeseries and anatomical images, as well as the statistical maps and questionnaire metadata. All the data is

[1] `http://www.logilab.fr/`

stored in a CubicWeb database[2], which makes it possible to perform advanced queries from a web interface or a programmatic language. In this thesis, we used other datasets from Neurospin, but they are not currently shared.

[2] http://www.cubicweb.org/

Resource name	Data type	# Tasks	URL
HCP	raw	7	https://db.humanconnectome.org
OpenfMRI	raw & maps	29	https://openfmri.org
fBIRN	raw	4	http://fbirnbdr.nbirn.net:8080/BDR/
NeuroVault	maps	92	http://neurovault.org/
BrainOmics	raw & maps	1	http://brainomics.cea.fr/localizer
BioMag 2010	raw	1	ftp://ftp.mrc-cbu.cam.ac.uk/personal/rik.henson/wakemandg_hensonrn/

Table 3.1: Table listing open fMRI resources. The number of tasks is relevant to indicate the variety of experimental designs even within a single study such as the HCP.

3.2 From diverse data sources to curated brain maps

From a technical standpoint, the main hindrance to sharing task fMRI is the lack of a common representation of the complex associated metadata. Task fMRI manipulates experimental conditions to study different mental processes. The analysis therefore requires the order and timing of the stimuli presentation during the acquisition, in addition to other MRI acquisition parameters. The analysis of large volumes of fMRI data presents additional challenges such as: i) homogenizing the processing streams, which otherwise may render studies incomparable [5], and ii) curating the data in an automatic or semi-automatic way.

In this section we present the steps we took to ready our database for analysis once the data was downloaded. We do not attempt to solve these issues in a general way, but rather take the approach of identifying the roadblocks, and finding pratical solutions for our project. Finally, we briefly describe all the studies we use in this thesis.

3.2.1 Organization

The neuroimaging community recognizes that the lack of metadata standards hurts the sharing and reuse of data, as well as the reproducibility of science [36]. Such standards are currently being developed [3] [13], and extend previous work [11]. To be useful, they will however first have to be supported by popular neuromaging software packages such as SPM and FSL, and such functionality is under way [24]. A wide adoption would ease the integration of datasets in the future. In the meantime, resources provide documentation or follow arbitrary standards derived from software packages that makes it possible to merge several datasets in a common organization.

[3] http://nidm.nidash.org/

OpenfMRI We chose to follow the OpenfMRI file layout, as the majority of our data came from there, and its organization is for our work straightforward to customize if needed. OpenfMRI's organization is mainly derived from FSL, and is described in Figure 3.4. The directory structure represents

the usual information regarding subjects, fMRI tasks and runs, and additional MRI modalities such as T1-weigthed images. More importantly, it stores the metadata related to the experimental design in a set of files that eases the process of scripting the analysis of the data. The experimental conditions and the associated contrasts are respectively stored in the "condition_key.txt" and "task_contrast.txt" under the model directory. The onsets specifying the stimuli presentation are stored in column formated files containing the timing, duration and weights for each stimulus presentation. All across the structure, files, subjects and conditions are named with generic labels such as "sub00x", "model00x", "cond00x". The mapping to human readable labels is located in relevant "key.txt" files. This approach may seem counter intuitive at first glance, but renders the scripting trivial and robust as **file and directory names are always the same across datasets**. Finally, even though the directory hierarchy allows to store raw and processed data within the same study directory, we chose to replicate the hierarchy in separate directories for each processing step. This is a data management issue, and enables to quickly erase and re-process some data if something went afoul along the processing stream.

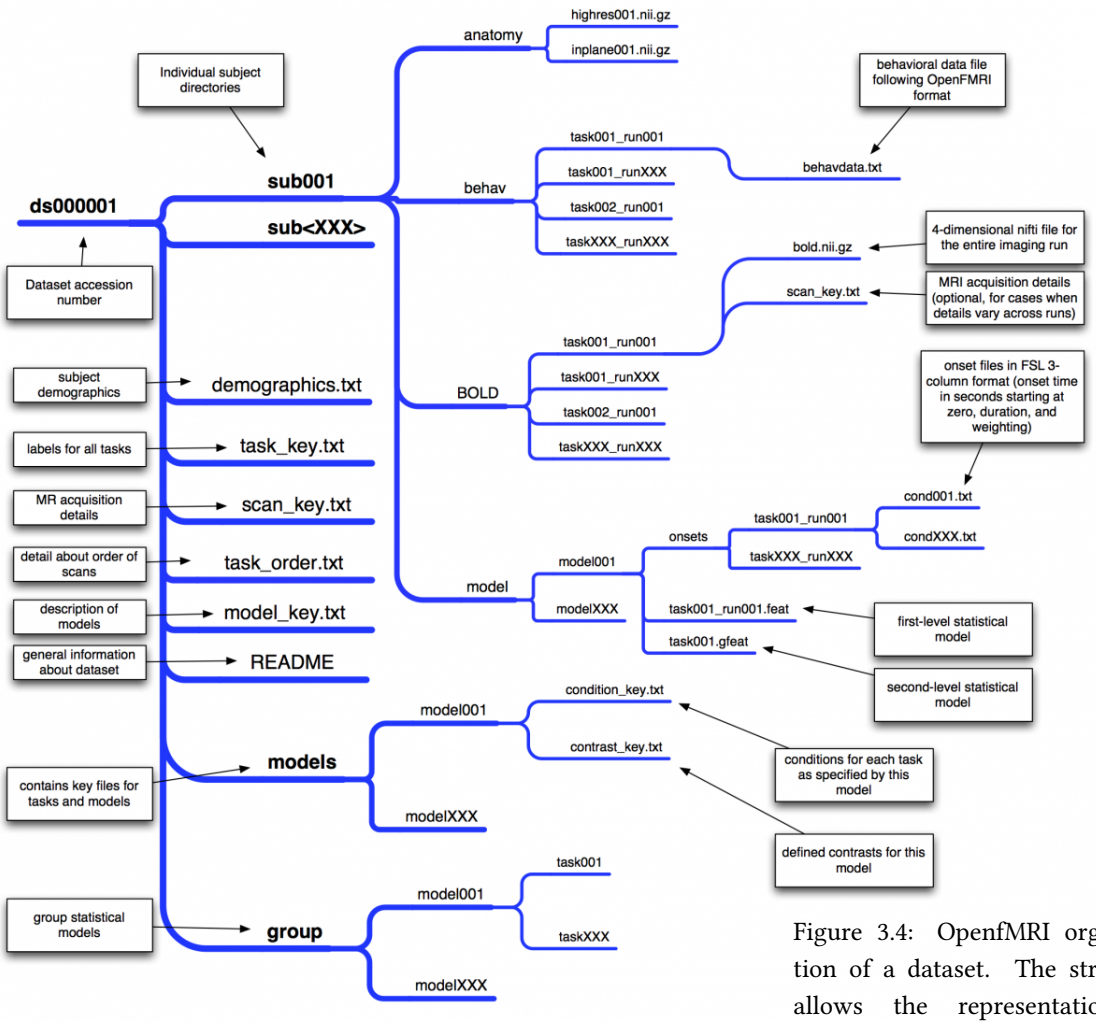


Figure 3.4: OpenfMRI organization of a dataset. The structure allows the representation of subjects, tasks, runs, models, as well as the experimental design. SOURCE: `https://openfmri.org/content/data-organization`

SPM SPM is a software for statistical processing of brain maps that Neurospin researchers use routinely to analyze their fMRI data. SPM stores all

the metadata related to the experimental design in the "SPM.mat" file, which is just a dump of an internal SPM data structure. Documentation may be found in SPM's code, or on webpages of thorough users who reported their findings [4]. The variables necessary to extract the design from an experiment are listed in Table 3.2. Additional variables contain the paths from the computed contrast maps, and from the pre-processed timeseries used as input data for the GLM. Similarly, other .mat files contain information on pre-processing and paths of raw timeseries and anatomy images. Scipy includes utilities to read such files from Matlab. We developed Python code available on Github[5] to help parsing and dumping SPM data in a new organization. It is not anywhere near production ready, but may serve as a prototype before integration in a library such as NiPy or Nilearn.

[4] http://www.its.caltech.edu/~nsulliva/spmdatastructure.htm

[5] https://github.com/schwarty/load_data

Variable name	Description
SPM.xY.RT	Repetition time.
SPM.[Sess].[U].name	Name of experimental condition.
SPM.[Sess].[U].ons	Condition trials onsets.
SPM.[Sess].[U].dur	Condition trials durations.
SPM.[xCon].name	Contrast name.
SPM.[xCon].c	Contrast weights (vector).
SPM.xX.X	Design matrix.
SPM.xX.name	Names of design matrix regressors.

Table 3.2: Main variables from SPM.mat to extract the design of an experiment. Square brackets represent list elements. Additional fields include contrasts and timeseries paths.

HCP The HCP provides a detailed documentation of its files structure online [6]. Roughly, the data are split in different directories between raw and pre-processed, and then in sub-directories for tasks. The design information is also stored following the FSL format and it is extremely straightforward to convert this organization in the one of OpenfMRI.

[6] http://www.humanconnectome.org/documentation/S500/HCP_S500+MEG2_Release_Appendix_III.pdf

fBIRN fBIRN contains data from schizophrenic patients as well as healthy subjects. The metadata regarding subjetcs are stored in a myriad of CSV and Excel spreadsheets. The experimental design is stored in files from the E-Prime software [7]. The use of proprietary formats makes it necessary to develop ad-hoc scripts to organize and analyze the data.

[7] http://www.pstnet.com/

3.2.2 Processing

The analysis of fMRI data requires a complex pre-processing stream that accounts for specificities of the acquisition sequences. Moreover, it enables normalizing the subjects data into a common brain space to make them comparable. Several softwares and steps are possible, and in the context of multi-study analysis, it is a considerable advantage to control all the parameters in order to maintain comparability. Reports indicate that varying pre-processing parameters may change significantly the results [40, 5].

fMRI Pre-processing

The usual fMRI pre-processing steps include slice timing correction, motion correction and spatial realignment, coregistration of fMRI and anatomical images, spatial normalization, and finally spatial smoothing. We use a custom pipeline available in pypreprocess [8] that relies on SPM for the actual processing, and Nipype [14] for the Python interfaces. Pypreprocess includes an example script "openfmri_preproc.py", that serves as a command line interface to process any dataset following the OpenfMRI data specification. It is also able to fetch the data if the dataset is hosted on the OpenfMRI repository and a valid identifier is given.

[8] https://github.com/neurospin/pypreprocess

Statistical Modeling

The standardized organization similarly enables to automate the generation of the experimental design. We use NiPy [9] [27] to perform the fMRI first-level GLM, as this library nicely integrates with the rest of our Python software stack. We initially attempted to use SPM to compute the statistical maps, but it estimates a separate mask for each subject and replaces all missing values by NaNs (Not a Number), which is extremely inconvenient for subsequent analyses. A possible but unpratical solution is to extrapolate missing values as an additional step. NiPy however makes it trivial to use an arbitrary mask. We created a mask from the tissue probability maps from SPM, and only kept voxels that have a 30 percent probability or more to lie in the grey matter across subjects.

[9] https://github.com/nipy/nipy

The additional difficulty in the statistical modeling of a large volume of data, is the generation of the statistical models themselves. Most studies model experimental conditions as separate regressors in the design matrix. In some cases, an event never happens and results in a null regressor that should not be passed to the model. For example, a study that models the correct and incorrect responses from subjects might have no incorrect, or no correct responses from some subjects. Contrasts are encoded as vectors, and are assumed to be identical for all the subjects. The modeling scripts have therefore to take into account that the contrast length might vary across the subjects.

Quality control

Quality control is essential to the analysis of fMRI data, even more so in the context analysing multiple datasets coming from variate sources. This process should ideally be fully automated, or to a lesser extent provide ways to quickly assess the quality and state of the processed data. We adopted 2 strategies in this thesis: i) the generation of graphical reports of the preprocessing, and ii) a heuristic to detect major outliers. These strategies eliminates most of the obvious problems that arise during the analysis stages, but may fall short in some cases as it relies on human input.

The first strategy relies on reports generated by Pypreprocess, that show the results of the SPM pre-processing. The reports display graphically subjects' movement during acquisition, as well as the results of normalization and co-registation. The most common issue comes from the NIfTI format headers that encode brain orientation. This information tries to alleviate

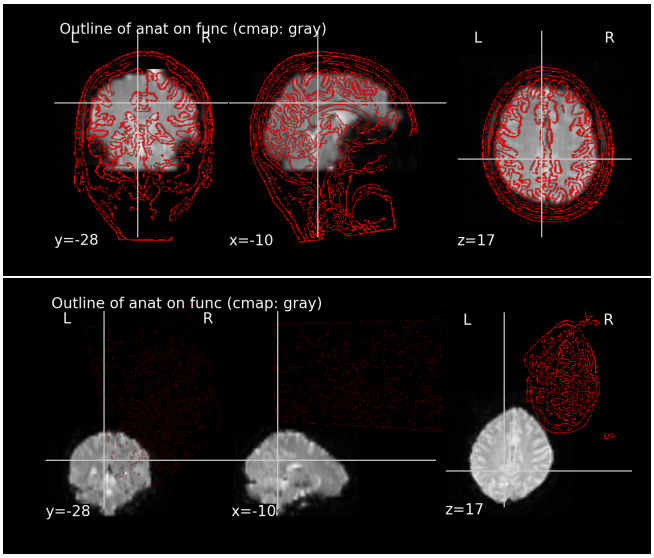


Figure 3.5: The outline of functional data with the corresponding subject's anatomy in background. The top image shows a valid preprocessing whereas the bottom one shows a preprocessing that failed because of faulty information in the NIfTI headers.

.

problems regarding left-right brain orientation, but is not consistently informed and interpreted by the neuroimaging softwares. Figure 3.5 shows the typical registation report for 2 subjects, one of which failed due to faulty NIfTI headers. The pratical solution to this problem is to first use the headers, and if it does not work remove them with the "fslorient" tool from FSL. We initially intented to use the data from fBIRN, but its processing failed inconsistently with or without the headers making their use all but impossible.

The second strategy to automatically detect major outliers was applied on statistical maps. It is often easy, given a set of maps and knowing what they are about to "see" if some are drastically different from the others. From that observation we developed a heuristic that computes a robust mean from the maps of a task on the voxel-level and exclude those that deviate too much from that mean. Figure 3.6 illustrates this by showing (from left to right) the mean maps of an auditory task, a correct map from one subject, and an incorrect one from another subject's same task. This is an example of outlier that this heuristic captures.

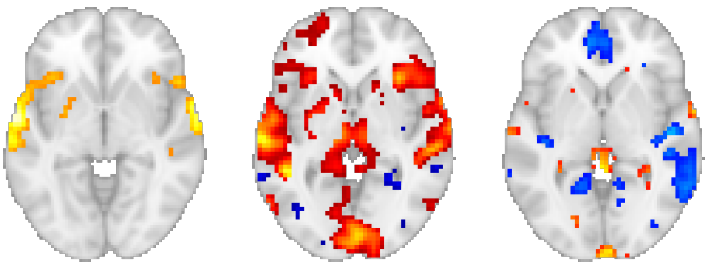


Figure 3.6: From left to right: group mean of an autitory task, correct statistical map of one subject, incorrect map of another subject. Automated methods are able to capture this kind of outliers.

.

3.2.3 Datasets

This section presents an overview of the database used in this thesis. A more detailed description of the datasets is available in Appendix A. Tables 3.3, 3.4, 3.5, and 3.6 provide a more compact listing of tasks and studies. Note that the dataset from BioMag's data contest has recently been uploaded to OpenfMRI and is accessible under the ds117 identifier. As as whole, the database accounts for 30 studies, 50 tasks, over 800 subjects, and 7K statistical maps. Considering all the raw and processed data amounts to close to 1TB of disk space, without the backups. This is also without considering the actual exploitation of the database which creates even more data. Figure 3.8 shows the distribution of tasks, and MRI scanners in the database depending on the orginal data source. This shows that i) Neurospin is our primary source of data and mainly from one MRI scanner, ii) OpenfMRI is our second largest source of data and is probably more challenging to analyze due to its scanner variability. We did not use the relational and social tasks from the HCP.

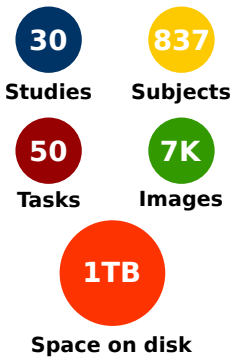


Figure 3.7: Database in figures.

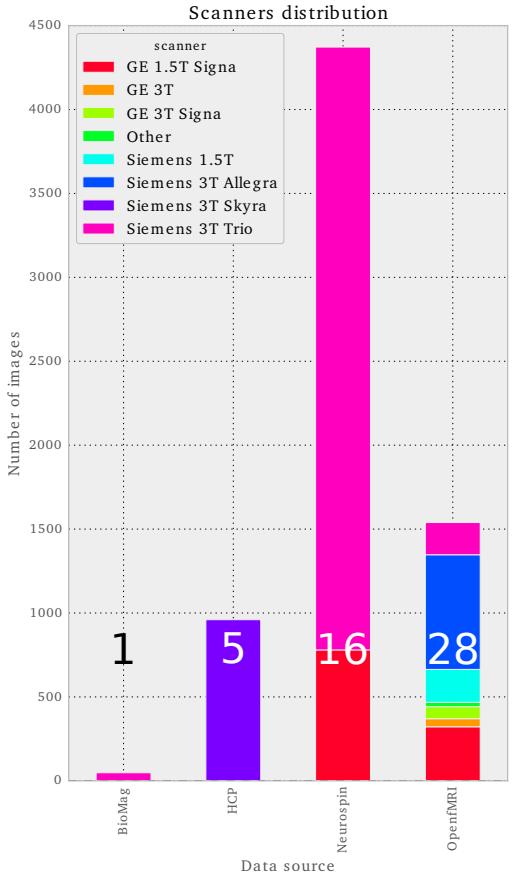


Figure 3.8: Distribution of scanners for the different data sources. The numbers inside each bar indicate the amount of unique fMRI tasks per data source.

.

3.3 Conlusion

Scaling up from individual studies leverages the recent efforts of the community on data sharing and standardization. We demonstrate that a common data organization is key to automate data integration and processing streams, and commonly adopted standards would further ease the process. To our knowledge, no standard or ad-hoc format is currently attempting to represent quality information. This is a tricky problem, as data usability is dependent on the intended use. In our opinion, quality assessment formats and automated methods, would however help to solve the main bottleneck to setting up a large fMRI database, as it still requires a lot of manual work. Contrary to popular belief, processing time and the associated processing power is not a problem: modern workstations are able to process the described database in a matter of days. The curation however takes a lot longer, and necessitates to adapt parameters, and re-run datasets multiple times. In the end, more than a universal standard, what we need to scale up are documented formats, and automated methods.

Bibliography

[1] Adam R Aron, Mark A Gluck, and Russell A Poldrack. Long-term test–retest reliability of functional mri in a classification learning task. *Neuroimage*, 29(3):1000–1006, 2006.

[2] Adam R Aron, Tim E Behrens, Steve Smith, Michael J Frank, and Russell A Poldrack. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (mri) and functional mri. *The Journal of Neuroscience*, 27(14):3743–3752, 2007.

[3] Jeffrey R. Binder, William L. Gross, Jane B. Allendorfer, Leonardo Bonilha, Jessica Chapin, Jonathan C. Edwards, Thomas J. Grabowski, John T. Langfitt, David W. Loring, Mark J. Lowe, Katherine Koenig, Paul S. Morgan, Jeffrey G. Ojemann, Christopher Rorden, Jerzy P. Szaflarski, Madalina E. Tivarus, and Kurt E. Weaver. Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *NeuroImage*, 54(2):1465–1475, 2011.

[4] B. B. Biswal, M. Mennes, X. N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, et al. Toward discovery science of human brain function. *Proc Ntl Acad Sci*, 107:4734, 2010.

[5] Joshua Carp. On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. *Frontiers in neuroscience*, 6, 2012.

[6] Fulvia Castelli, Francesca Happé, Uta Frith, and Chris Frith. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, 12(3):314–325, 2000.

[7] Mauricio R Delgado, Leigh E Nystrom, C Fissell, DC Noll, and Julie A Fiez. Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of neurophysiology*, 84(6):3072–3077, 2000.

[8] Keith J Duncan, Chotiga Pattamadilok, Iris Knierim, and Joseph T Devlin. Consistency and variability in functional localisers. *Neuroimage*, 46(4):1018–1026, 2009.

[9] Roy Thomas Fielding. *Architectural styles and the design of network-based software architectures*. PhD thesis, University of California, Irvine, 2000.

[10] Karin Foerde, Barbara J Knowlton, and Russell A Poldrack. Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences*, 103(31):11778–11783, 2006.

[11] Syam Gadde, Nicole Aucoin, Jeffrey S Grethe, David B Keator, Daniel S Marcus, and Steve Pieper. Xcede: an extensible schema for biomedical data. *Neuroinformatics*, 10(1):19–32, 2012.

[12] Baptiste Gauthier, Evelyn Eger, Guido Hesselmann, Anne-Lise Giraud, and Andreas Kleinschmidt. Temporal tuning properties along the human ventral visual stream. *The Journal of Neuroscience*, 32(41):14433–14441, 2012.

[13] Satrajit Ghosh, Nolan Nichols, Syam Gadde, Jason Steffener, and David Keator. Xcede-dm: A neuroimaging extension to the w3c provenance data model. *Frontiers in Neuroinformatics*, (7), 2012. ISSN 1662-5196. doi: 10.3389/conf.fninf.2013.08.00007.

[14] Krzysztof Gorgolewski, Christopher D Burns, Cindee Madison, Dav Clark, Yaroslav O Halchenko, Michael L Waskom, and Satrajit S Ghosh. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in neuroinformatics*, 5, 2011.

[15] Krzysztof J Gorgolewski, Amos Storkey, Mark E Bastin, Ian R Whittle, Joanna M Wardlaw, and Cyril R Pernet. A test-retest fMRI dataset for motor, language and spatial attention functions. *GigaScience*, 2(1):6, 2013.

[16] Krzysztof Jacek Gorgolewski, Gael Varoquaux, Gabriel Rivera, Yannick Schwartz, Satrajit S. Ghosh, Camille Maumet, Thomas E. Nichols, Russell A. Poldrack, Jean-Baptiste Poline, Tal Yarkoni, and Daniel S. Margulies. Neurovault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *bioRxiv*, 2014. doi: 10.1101/010348.

[17] JV Haxby, IM Gobbini, ML Furey, A Ishai, JL Schouten, and P Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293:2425, 2001.

[18] Richard N Henson, Daniel G Wakeman, Vladimir Litvak, and Karl J Friston. A parametric empirical bayesian framework for the EEG/MEG inverse problem: generative models for multi-subject and multi-modal integration. *Frontiers in human neuroscience*, 5, 2011.

[19] Koji Jimura, Fabienne Cazalis, Elena RS Stover, and Russell A Poldrack. The neural basis of task switching changes with skill acquisition. *Frontiers in Human Neuroscience*, 8, 2014.

[20] David B Keator, Jeffrey S Grethe, D Marcus, B Ozyurt, Syam Gadde, Sean Murphy, Steve Pieper, D Greve, R Notestine, H Jeremy Bockholt, et al. A national human neuroimaging collaboratory enabled by the biomedical informatics research network (birn). *Information Technology in Biomedicine, IEEE Transactions on*, 12(2):162–172, 2008.

[21] AM Kelly, Lucina Q Uddin, Bharat B Biswal, F Xavier Castellanos, and Michael P Milham. Competition between functional brain networks mediates behavioral variability. *Neuroimage*, 39(1):527–537, 2008.

[22] André Knops, Bertrand Thirion, Edward M Hubbard, Vincent Michel, and Stanislas Dehaene. Recruitment of an area involved in eye movements during mental arithmetic. *Science*, 324(5934):1583–1585, 2009.

[23] Stephen Manuck, Sarah Brown, Erika Forbes, and Ahmad Hariri. Temporal stability of individual differences in amygdala reactivity. *American Journal of Psychiatry*, 164(10):1613–1614, 2007.

[24] Camille Maumet, Guillaume Flandin, B Nolan Nichols, Jason Steffener, Karl Helmer, Krzysztof J. Gorgolewski, Tibor Auer, Gully Burns, Fariba Fana, Jessica A. Turner, Thomas E. Nichols, Satrajit S Ghosh, Jean-Baptiste Poline, and David B Keator. Extending ni-dm to share the results and provenance of a neuroimaging study: implementation within spm and fsl. *Frontiers in Neuroinformatics*, (31), 2014. ISSN 1662-5196. doi: 10.3389/conf.fninf.2014.18.00031.

[25] Vincent Michel, Yannick Schwartz, Philippe Pinel, Olivier Cayrol, Antonio Moreno, Jean-Baptiste Poline, Vincent Frouin, and Dimitri Papadopoulos Orfanos. Brainomics: A management system for exploring and merging heterogeneous brain mapping data. *OHBM 2013 19th Annual Meeting of the Organization for Human Brain Mapping*, 2013.

[26] Michael B Miller, Christa-Lynn Donovan, John D Van Horn, Elaine German, Peter Sokol-Hessner, and George L Wolford. Unique and persistent individual patterns of brain activity across different memory retrieval tasks. *Neuroimage*, 48(3):625–635, 2009.

[27] K Jarrod Millman and Matthew Brett. Analysis of functional magnetic resonance imaging in python. *Computing in Science & Engineering*, 9(3):52–55, 2007.

[28] Joseph M Moran, Eshin Jolly, and Jason P Mitchell. Social-cognitive deficits in normal aging. *The Journal of Neuroscience*, 32(16):5553–5561, 2012.

[29] Takato Morioka, Tomoya Yamamoto, Akira Mizushima, Shozo Tombimatsu, Hiroshi Shigeto, Kanehiro Hasuo, Shunji Nishio, Kiyotaka Fujii, and Masashi Fukui. Comparison of magnetoencephalography, functional mri, and motor evoked potentials in the localization of the sensory-motor cortex. *Neurological research*, 17(5):361–367, 1995.

[30] Christophe Pallier, Anne-Dominique Devauchelle, and Stanislas Dehaene. Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, 108(6):2522–2527, 2011.

[31] Philippe Pinel and Stanislas Dehaene. Genetic and environmental contributions to brain activation during calculation. *NeuroImage*, 81:306–316, 2013.

[32] Philippe Pinel, Bertrand Thirion, Sébastien Meriaux, Antoinette Jobert, Julien Serres, Denis Le Bihan, Jean-Baptiste Poline, and Stanislas Dehaene. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC neuroscience*, 8(1):91, 2007.

[33] Russell A Poldrack and Krzysztof J Gorgolewski. Making big data open: data sharing in neuroimaging. *Nature neuroscience*, 17(11):1510–1517, 2014.

[34] Russell A Poldrack, J Clark, EJ Pare-Blagoev, D Shohamy, J Creso Moyano, C Myers, and MA Gluck. Interactive memory systems in the human brain. *Nature*, 414(6863):546–550, 2001.

[35] Russell A Poldrack, Deanna M Barch, Jason P Mitchell, Tor D Wager, Anthony D Wagner, Joseph T Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael P Milham. Toward open sharing of task-based fMRI data: the openfmri project. *Frontiers in neuroinformatics*, 7, 2013.

[36] Jean-Baptiste Poline, Janis L Breeze, Satrajit Ghosh, Krzysztof Gorgolewski, Yaroslav O Halchenko, Michael Hanke, Christian Haselgrove, Karl G Helmer, David B Keator, Daniel S Marcus, et al. Data sharing in neuroimaging research. *Frontiers in neuroinformatics*, 6, 2012.

[37] Tom Schonberg, Craig R Fox, Jeanette A Mumford, Eliza Congdon, Christopher Trepel, and Russell A Poldrack. Decreasing ventromedial prefrontal cortex activity during sequential risk-taking: an fMRI investigation of the balloon analog risk task. *Frontiers in neuroscience*, 6, 2012.

[38] G Schumann, E Loth, T Banaschewski, A Barbot, G Barker, C Büchel, PJ Conrod, JW Dalley, H Flor, J Gallinat, et al. The imagen study: reinforcement-related behaviour in normal brain function and psychopathology. *Molecular psychiatry*, 15(12):1128–1139, 2010.

[39] Rachelle Smith, Kamyar Keramatian, and Kalina Christoff. Localizing the rostrolateral prefrontal cortex at the individual level. *Neuroimage*, 36(4):1387–1396, 2007.

[40] Stephen Strother, Stephen La Conte, Lars Kai Hansen, Jon Anderson, Jin Zhang, Sujit Pulapura, and David Rottenberg. Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. a preliminary group analysis. *Neuroimage*, 23:S196–S207, 2004.

[41] Bertrand Thirion, Gaël Varoquaux, Olivier Grisel, Cyril Poupon, and Philippe Pinel. Principal component regression predicts functional responses across individuals. pages 741–748, 2014.

[42] Sabrina M Tom, Craig R Fox, Christopher Trepel, and Russell A Poldrack. The neural basis of loss aversion in decision-making under risk. *Science*, 315(5811):515–518, 2007.

[43] Melina R Uncapher, J Benjamin Hutchinson, and Anthony D Wagner. Dissociable effects of top-down and bottom-up attention during episodic encoding. *The Journal of Neuroscience*, 31(35):12613–12628, 2011.

[44] Laurianne Vagharchakian, Ghislaine Dehaene-Lambertz, Christophe Pallier, and Stanislas Dehaene. A temporal bottleneck in the language comprehension network. *The Journal of Neuroscience*, 32(26):9089–9102, 2012.

[45] David C Van Essen, Kamil Ugurbil, E Auerbach, D Barch, TEJ Behrens, R Bucholz, A Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

[46] John D Van Horn, Jeffrey S Grethe, Peter Kostelec, Jeffrey B Woodward, Javed A Aslam, Daniela Rus, Daniel Rockmore, and Michael S Gazzaniga. The functional magnetic resonance imaging data center (fmridc): the challenges and rewards of large–scale databasing of neuroimaging studies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1412):1323–1339, 2001.

[47] Tor D Wager, Matthew L Davidson, Brent L Hughes, Martin A Lindquist, and Kevin N Ochsner. Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron*, 59(6):1037–1050, 2008.

[48] Gui Xue and R Poldrack. The neural substrates of visual perceptual learning of words: implications for the visual word form area hypothesis. *Cognitive Neuroscience, Journal of*, 19(10):1643–1655, 2007.

[49] Gui Xue, Adam R Aron, and Russell A Poldrack. Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex*, 18(8):1923–1932, 2008.

Table 3.3: List of the datasets and associated tasks from OpenfMRI. The dataset # is its identifier in this thesis.

Dataset #	Accession #	Task #	Task description	References #
1	ds001	1	Balloon analog risk task	Schonberg et al. [37]
2	ds002	1	Probabilistic classification	Aron et al. [1]
3	ds002	2	Deterministic classification	Aron et al. [1]
4	ds002	3	Mixed event-related probe	Aron et al. [1]
5	ds003	1	Rhyme judgment	Xue and Poldrack [48]
6	ds005	1	Mixed-gambles task	Tom et al. [42]
7	ds006A	1	Living-nonliving decision with plain or mirror-reversed text	Jimura et al. [19]
8	ds007	1	Stop signal with manual response	Xue et al. [49]
9	ds007	2	Stop signal with letter naming	Xue et al. [49]
10	ds007	3	Stop signal with pseudoword naming	Xue et al. [49]
11	ds008	1	Stop signal	Aron et al. [2]
12	ds008	2	Conditional stop signal	Aron et al. [2]
13	ds009	1	Balloon analog risk task	Cohen and Poldrack, unpublished
14	ds009	2	Stop signal	Cohen and Poldrack, unpublished
15	ds009	3	Emotional regulation	Cohen and Poldrack, unpublished
16	ds009	4	Discounting	Cohen and Poldrack, unpublished
17	ds011	1	Tone-counting	Foerde et al. [10]
18	ds011	2	Single-task weather prediction	Foerde et al. [10]
19	ds011	3	Dual-task weather prediction	Foerde et al. [10]
20	ds011	4	Classification probe without feedback	Foerde et al. [10]
21	ds017A	1	Probabilistic classification	Rizk-Jackson et al., unpublished
22	ds017A	2	Selective stop-signal task	Rizk-Jackson et al., unpublished
23	ds051	1	Abstract-concrete judgment	Alvarez and Poldrack, unpublished
24	ds052	1	Weather prediction	Poldrack et al. [34]
25	ds052	2	Reversal weather prediction	Poldrack et al. [34]
26	ds101	1	Simon task	Kelley and Milham, unpublished
27	ds102	1	Flanker task	Kelly et al. [21]
28	ds105	1	Object viewing	Haxby et al. [17]
29	ds107	1	One-back task	Duncan et al. [8]
30	ds108	1	Emotion regulation	Wager et al. [47]
31	ds109	1	Theory of mind with manual response	Moran et al. [28]
32	ds110	1	Incidental encoding task using Posner cueing paradigm with object v greeble judgment	Uncapher et al. [43]
33	ds114	1	Overt word repetition	Gorgolewski et al. [15]
34	ds114	2	Covert verb generation	Gorgolewski et al. [15]
35	ds114	3	Finger foot lips	Gorgolewski et al. [15]
36	ds114	4	Overt verb generation	Gorgolewski et al. [15]
37	ds114	5	Line bisection	Gorgolewski et al. [15]

Table 3.4: List of the datasets and associated tasks from Neurospin. The dataset # is its identifier in this thesis.

Dataset #	Accession #	Task #	Task description	References #
38	amalric2012mathematicians	1	Visual recognition task (one-back)	Amalric et al., unpublished
39	amalric2012mathematicians	2	Localizer task	Amalric et al., unpublished
40	cauvet2009muslang	1	Music task	Cauvet et al., unpublished
41	cauvet2009muslang	2	Language task	Cauvet et al., unpublished
42	devauchelle2009sentence	1	Language task	Pallier et al. [30]
43	gauthier2009resonance	1	Continuous face house block	Gauthier et al. [12]
44	gauthier2009resonance	2	Discontinuous face house block 400ms frequency	Gauthier et al. [12]
45	gauthier2009resonance	3	Discontinuous face house block 800ms frequency	Gauthier et al. [12]
46	gauthier2009resonance	4	Object localizer	Gauthier et al. [12]
47	gauthier2010resonance	1	Continuous face house block	Gauthier et al. [12]
48	gauthier2010resonance	2	Continuous face house block with distractor	Gauthier et al. [12]
49	gauthier2010resonance	3	Object localizer	Gauthier et al. [12]
50	knops2009recruitment	1	Calculation task	Knops et al. [22]
51	knops2009recruitment	2	Saccades task	Knops et al. [22]
52	knops2009recruitment	3	Saccades localizer	Knops et al. [22]
53	pinel2007fast	1	Localizer task	Pinel et al. [32]
54	pinel2009twins	1	Object recognition task	Pinel and Dehaene [31]
55	pinel2009twins	2	Arithmetics and saccades task	Pinel and Dehaene [31]
56	pinel2009twins	3	Language task	Pinel and Dehaene [31]
57	pinel2012archi	1	Localizer task	Thirion et al. [41]
58	pinel2012archi	2	Social task	Thirion et al. [41]
59	pinel2012archi	3	Emotional task	Thirion et al. [41]
60	pinel2012archi	4	Parietal task	Thirion et al. [41]
61	vagharchakian2012temporal	1	Visual language compression task	Vagharchakian et al. [44]
62	vagharchakian2012temporal	2	Auditory language compression task	Vagharchakian et al. [44]

Table 3.5: List of the tasks from the HCP. The dataset # is its identifier in this thesis.

Dataset #	Accession #	Task #	Task description	References #
63	HCP	1	Emotion task	Manuck et al. [23]
64	HCP	2	Gambling task	Delgado et al. [7]
65	HCP	3	Language task	Binder et al. [3]
66	HCP	4	Motor task	Morioka et al. [29]
67	HCP	5	Relational task	Smith et al. [39]
68	HCP	6	Social task	Castelli et al. [6]
69	HCP	7	Working memory task	Miller et al. [26]

Table 3.6: Task from the BioMag 2010 dataset.

Dataset #	Accession #	Task #	Task description	References #
70	henson2010faces	1	Face recognition task	Henson et al. [18]

4 Functional localization by meta-analysis

Chapter 4 investigates the value of using a meta-analytic database to create **functional localizers**. As it was still in the early stages of its making, we only use here a small fraction of the database described in the previous chapter.

In this chapter we describe two concomitant contributions. Both contributions employ what we call **transfer learning** [16], but is transfer learning in its simplest form and usually referred to as **classifier generalization**. We compare the generalization of classification on functional tasks, *i.e.* training a classifier on task A and without further training predicting task B, to the performance of a classifier in the usual classification setting, *i.e.* training a classifier on task A and predicting on the same task within a cross validation loop. To distinguish this usual procedure to classifier generalization we refer to it as **inline learning**. Finally the contributions designates the tasks as *source task* and *target task* to indicate the direction of the generalization. In this introduction we simply call them task A and B.

The first contribution explores the ability of a classifier trained on task A to generalize on task B, and use its predictive features to **better condition hypothesis testing** on task B. We do not attempt to use generalization blindly, i.e. try all possible task combinations in the database. We perform generalization only when we assume that both tasks have something in common, so that the generalization makes sense. The tasks ideally –but not necessarily– come from separate studies. The gist of this method is to use classifier generalization to validate the use of a candidate task A, to define ROIs to analyse task B. Figure 4.1 depicts the classification procedures, where task A is a French and Korean auditory task, and task B a French and jabberwocky (pseudowords) visual tasks. The similarity in both tasks is the comprehension versus incomprehension of a stimulus. Inline learning is here merely performed to have a comparison reference for the generalized classifier. We use the classifier's predictive voxels to define regions of interest. The subsequent analyses, show that we **increase statistical power** by using these regions. We also show that a similar approach based on databases of activation coordinates would be less successful. We therefore demonstrate the validity of this approach to learn a functional localizer with a meta-analytic database.

The second contribution aims to solve **the threshold selection limi-**

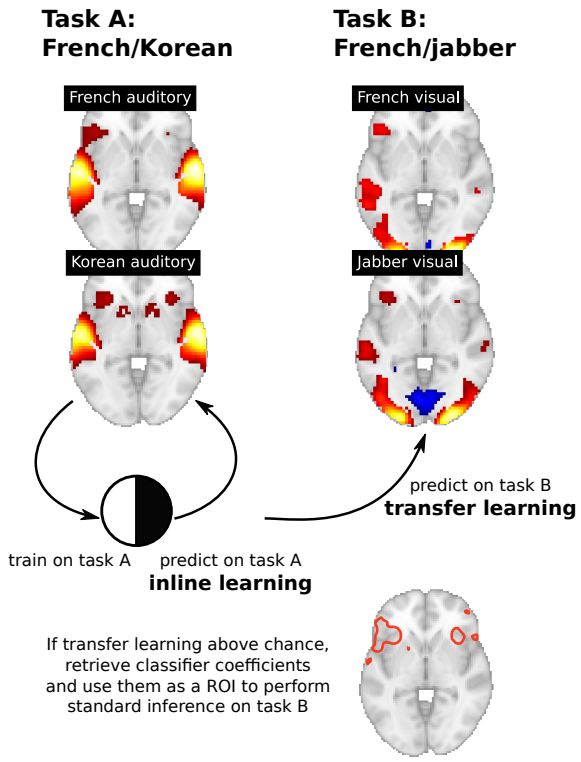


Figure 4.1: Classifier generalization (transfer learning) **validates** the use of task A to define regions of interest on task B. Regions of interest are selected by thresholding the classifier's coefficient to keep a fraction of 5% of the brain. Subsequent standard **inference analysis** on task B with the regions of interest shows inscreased statistical power.

tation from contribution one. Here, we slightly change the way to select features. We use a linear model with ℓ_2 penalization, but perform an univariate screening first. We increase the amout of features retained by the univariate selection, until we reach the full brain. At each scale, we compare the classification scores of the generalized classifier (trained on task A, predicted on task B), and the inline classifier (trained on task B, predicted on task B) which uses the same feature selection procedure. The intention is to select the scale at which both classifications yield the same performance. We will show that it is difficult to select a scale with this approach, and that the resulting regions do not delineate specific brain regions. This procedure is illustrated in Figure 4.2.

To overcome this difficuly we introduce a novel approach that we call **selection transfer**. This procedure does not attempt to generalize a classifier. Instead, it selects regions from task A with the univariate screening, and builds a predictive model on task B. As shown in Figure 4.3. It compares the ability of this classifier to the inline learning procedure. This enables the prediction curve to converge and select a scale at which both classifiers performs equally well. This scale permit us to expose regions specific to both tasks. The procedure can be seen as a **multivariate alternative to conjunction**.

The contributions developed in this chapter have been published in:

- Y. Schwartz, G. Varoquaux, C. Pallier, P. Pinel, J.B. Poline, and B. Thirion, *Improving accuracy and power with transfer learning using a*

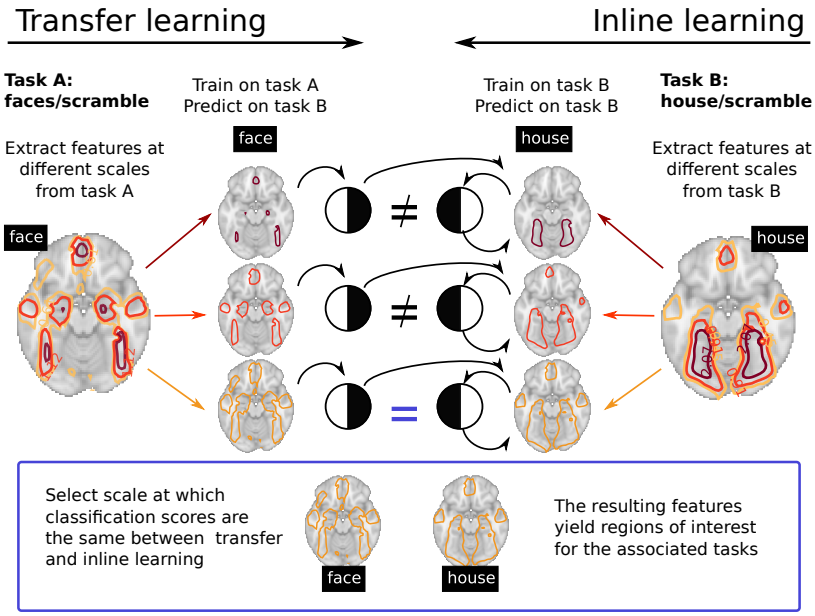


Figure 4.2: Use of **transfer learning** to select the scale at which its performance is similar to inline learning. Resulting regions are unspecific to the investigated tasks.

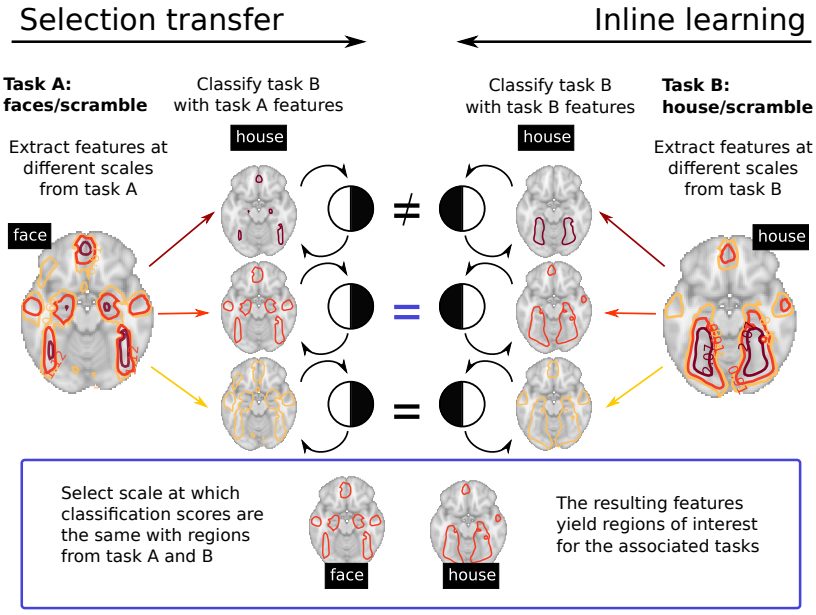


Figure 4.3: Use of **selection transfer** to select the scale at which its performance is similar to inline learning. Resulting regions are more specific to the investigated tasks.

meta-analytic database, MICCAI, 2012, pages 1–8.

- Y. Schwartz, G. Varoquaux, and B. Thirion, *On spatial selectivity and prediction across conditions with fMRI*, PRNI 2012 : 2nd International Workshop on Pattern Recognition in NeuroImaging, pages 53–56.

Contents

4.1 Improving accuracy and power with transfer learning using a meta-analytic database

4.1.1 Introduction

Multi-subject or multi-condition experiments are the workhorse of bio-medical imaging research, whether it be drug development or basic research. Imaging provides a wealth of information on the biomedical problem at hand. However the typical sample size is too small to fully exploit this information. For this reason, investigators often turn to previous studies in order to formulate hypotheses and restrict the search space, *i.e.* select a subset of anatomical or functional structures of interest to the current study. A typical case is that of early-stage clinical trials, for which the group size is very small, but that are most often based on previous results concerning the pathology under study. However, understanding the literature is increasingly difficult and requires a systematic approach, that takes the form of a *meta-analysis*, pooling results from multiple experiments that address a set of related research hypotheses [19].

In particular, brain imaging studies heavily rely on such meta-analyses [22], as the brain is still an ill-understood and complex organ. In functional Magnetic Resonance Imaging (fMRI) studies, typical group sizes range from 10 to 20 subjects, which is not always enough to warrant the reliability of brain-wide analysis [20]. More importantly, the subjects spend a finite time in the scanner, which limits the conditions under which a particual cognitive process is studied. For this reason, it is common practice to reduce the study to a set of regions of interest (ROIs) extracted from the literature. Investigators define these ROIs by extracting locations of peak activations from the literature [23], or from coordinate databases such as BrainMap [10]. While most of these meta-analyses are conducted on activation coordinates, the increase of data sharing opens the door to meta-analysis on full brain images which results in higher statistical power [18]. Previous statistical and modeling work on meta-analysis for fMRI has focused on better modeling of the reference database [22].

In this work, we are interested in the generalization power of meta-analyses on new data. We introduce a new meta-analysis method using a reference database of images to guide statistical analysis of a new dataset. In particular we rely on predictive models, useful to learn biomarkers, and use them to select relevant voxels in order to increase the statistical power of a new study.

4.1.2 Methods

Problem setting We start from a reference database made of l experiments, each comprising n^l contrasts acquired over multiple subjects. We denote the brain images by $\boldsymbol{X}^l \in \mathbb{R}^{n^l, p}$ with associated experimental condition $\boldsymbol{y}^l \in \{0, 1\}^{n^l}$. Given a new experiment, denoted *target*, $(\boldsymbol{X}^\star, \boldsymbol{y}^\star) \in (\mathbb{R}^{n^\star, p}, \{0, 1\}^{n^\star})$, we are interested in two problems: i) *(biomarkers)* can we predict $\boldsymbol{y}^\star$ from $\boldsymbol{X}^\star$? ii) *(inference)* can we test hypotheses on the links between $\boldsymbol{y}^\star$ and $\boldsymbol{X}^\star$, for instance in a linear model? These are ill-posed problems from the statistics standpoint, as $n^\star \ll p$. The root of the problem is

the dimensionality of the data: medical images are composed of many voxels, typically $p \approx 50\,000$ with fMRI. This large number of descriptors limits statistical inference power due to multiple testing; a problem that appears in predictive approaches as the curse of dimensionality. Here, we use our reference database to better condition this statistical problem.

Transfer learning The gist of our approach is to learn on some experiments of our database $(\boldsymbol{X}^l, \boldsymbol{y}^l)$ discriminative models that contain predictive information for the target experiment $(\boldsymbol{X}^\star, \boldsymbol{y}^\star)$. In machine learning, this problem is known as *transfer learning* [16]. The underlying assumption of transfer learning is the same as that for meta-analysis: the reference database should contain some common information with the target experiment. Here we use a simple form of transfer learning: we train a linear classifier on an experiment in the database that is similar from the neuroscientific point of view to the new data, and use it to predict the labels of the new data.

Selecting predictive features We use a sparse linear classifier, specifically an ℓ_1-penalized logistic regression. The motivation behind this choice of classifier is that it produces a sparse set of weights that can be used to select relevant voxels. In particular, under certain conditions, the classifier can recover with high probability the complete set of k features in $\boldsymbol{X}$ that are predictive of $\boldsymbol{y}$ for a sample size of $n_{\min} = O\left(k \log p\right)$ [1]. The logarithmic dependence in p is an appealing property in view of the dimensionality of medical imaging datasets.

In practical situations, it can be hard to control the errors on this feature selection, in particular as it depends on the choice of the amount of ℓ_1 penalty. For this reason, Meinshausen and Bühlmann [11] introduce randomized variants of sparse estimators, that can be seen as sampling the posterior probability of selection and keeping only features that are selected frequently. In particular, they establish non-asymptotic recovery results for the *randomized lasso*, which consists in applying the Lasso on random subsamples of the data and rescaling of the regressors. Here, we adapt this strategy to classification as the logistic regression is locally equivalent to a weighted least square and recovery results can carry from square-loss regression to logistic regression [1].

We want to use transfer learning to perform screening of the voxels, *i.e.* eliminate many voxels that are not related to our target experiment. For this purpose, we need a low probability of rejecting relevant variables. Each iteration of the sparse logistic regression in the randomized logistic can select reliably only $k_{\max} \approx n / \log p$ variables. In the worst case situation, we have k heavily-correlated variables and one of them is selected at random by the sparse logistic regression at each iteration. For each of these variables, the probability of selecting it less than s times during m iterations of the randomized logistic is given by the cumulative distribution function of a binomial with per trial success ratio $1/k$. If $s \leq m/k$, by Hoeffding's inequality, this probability goes to zero in $o\left(\exp m\right)$. In other words, if we impose a threshold $\tau = s/m$ on the selection frequency, we can recover a group of k correlated variables for $\tau \leq 1/k$.

Brain parcellations Although randomization relaxes the conditions on recovery, a remaining necessary condition is that the regressors of interest, *i.e.* the values $\boldsymbol{x}_i$ across the subjects of the k predictive voxels, must be weakly correlated[1]. Because of the large amount of smoothness present in medical images, in particular in group-level fMRI contrasts, these conditions cannot be satisfied. Indeed, values taken by a voxel are very similar to values taken by its neighbors. In addition, the numbers of subjects used in fMRI are often below the sample size required for good recovery. For these reasons we resort to feature agglomeration: using hierarchical clustering to merge neighboring voxels carrying similar information into parcels [12]. This strategy brings the double benefit of reducing the problem size, and thus the required sample size, and mitigating local correlation, at the expense of spatial resolution.

[1] Specifically, the condition for recovery with randomized lasso it is a lower bound on the conditioning of the sparse eigenvalues of the design matrix [11, theorem 2] and for sparse logistic regression the corresponding condition is a lower bound on the eigenvalues of the regressors of interest's covariance matrix [1, theorem 4].

4.1.3 Experiments and Results

FRMI datasets

We use 3 studies for this meta-analysis. The first study (*E1*) [17] is composed of 322 subjects and was designed to assess the inter-subject variability in some language, calculation, and sensorimotor tasks. The second study (*E2*) is similar to the first one in terms of stimuli, but its data was acquired on 35 pairs of twin-subjects. The last study (*E3*) [15] characterizes brain regions in charge of the syntactic and the semantic processing for the language. It was performed with 40 subjects, 20 of which were stimulated by pseudowords (jabberwocky stimuli) instead of actual meaningful sentences. We used in particular *E2* and *E3*, to learn regions for native versus non-native language comprehension (French versus Korean in *E2*), and better condition the French versus Jabberwocky experiments (*E3*). All the studies were pre-processed and analyzed with the standard fMRI analysis software SPM5. The data used for this work are the statistical images resulting from the intra-subject analyses across the 3 studies. *E1* has 34 contrasts images available, *E2* 56, and *E3* 28. The raw images were not always acquired on the same scanner. *E1* has data from a 3T SIEMENS Trio and a 3T Brucker scanner; *E2* data were acquired on a 1.5T GE Signa; and *E3* images come from the same 3T SIEMENS Trio.

Experimental results for prediction

Here we are interested in the *prediction* problem: using transfer learning to discriminate a pair of constrasts with an estimator trained on two other contrasts.

We used 4 different approaches to learn the discriminative models. The first approach relies on the activation likelihood estimate (ALE) method [9], as this is a commonly published method for coordinate-based meta-analyses. We extract the activation positions from the contrasts maps, and then apply a Gaussian kernel. We use the preferred FWHM of 10mm [21]. The other approaches directly use the contrast images. We name *raw contrasts* the method based on the contrasts voxels values; *contrast-specific parcels* the method that uses parcels from the training set: and *meta-analystic*

parcels the method that learns parcels from the full database. We evaluate on our base of contrasts the ability to do transfer learning, *i.e* to learn decision rules that carry over from one situation to another. Since we must make the assumption that the reference contrasts hold common information with the contrasts of interest, we do not try out all the possible combinations, but rather manually select pairs of contrasts from a single experiment that form a meaningful classification task (*e.g.*, computation versus reading, or Korean language versus French language). Out of all the possible combinations, we select 35 pairs of classification task, and subsequently combine them into 18 transfer pairs, on which it is reasonable to think that the transfer could occur (*e.g.*, computation and reading in visual instructions, transfer on computation and reading in auditory instructions). We first train a linear classifier within 6-fold cross validation test on a first set of pairs, setting the penalization amount by nested cross-validation, we call this step *inline learning*. We then re-use the discriminative model on a different pair of contrasts to perform the *transfer learning*. The 3 studies containing language related tasks, we can transfer between pairs within an experiment, and across experiments. Among the 18 selected transfer pairs, we find that 9 can give rise to such a transfer. Since a transfer is directed, we perform it both ways, which yields once again 18 transfer pairs to test upon. The associated prediction scores from the different methods are reported in Table 4.1. The general observation is that ALE yields a poorer prediction performance than any other method. This is true both for the transfer and inline predictions. We also find that brain parcellations scores similar to the raw contrasts images, and closer to the inline predictions. We find that while the contrast-specific parcels and meta-analytic parcels methods do not return the same parcels, they produce very close results. We can thus use the full database to learn a single reference parcellation to perform meta-analysis.

Names	Peaks		Contrasts		Parcels		Meta parcels	
	trans.	in.	trans.	in.	trans.	in.	trans.	in.
E1, comp./sent. → *E2*, comp./sent.	0.75	0.85	**0.88**	0.97	0.83	0.96	0.83	0.96
E2, comp./sent. → *E1*, comp./sent.	0.66	0.83	**0.88**	0.96	0.85	0.95	0.85	0.96
E3, jabb./French (L) → *E3*, jabb./French (S)	0.46	0.48	0.65	0.67	0.62	0.60	**0.67**	0.62
E3, jabb./French (S) → *E3*, jabb./French (L)	0.52	0.71	0.67	0.85	**0.71**	0.85	0.65	0.79
E3, jabb./French (L) → *E2*, Korean/French	0.65	0.46	0.73	0.79	0.65	0.81	**0.76**	0.85
E2, Korean/French → *E3*, jabb./French (L)	0.73	0.81	**0.79**	0.85	0.75	0.81	0.75	0.75

Table 4.1: Prediction scores for inline and transfer learning. trans.= transfer; in.= inline; comp.= computation, sent.= sentences (reading), jabb.= jabberwocky; S= sentence with one word constituents, L= one constituent long sentence.

Experimental results for inference

Here we are interested in the *inference* problem: using transfer learning to help hypothesis testing on a target dataset. In the following, we only consider a specific transfer, namely the last line in Table 4.1: we learn a model discriminating French native speakers reading French or Korean, and apply it on another experiment in which French subjects had to read French or jabberwocky. This transfer is interesting as it involves two different experiments acquired on different scanners, and cognitive paradigms that share a similar expression, incomprehension of language stimuli. As can be seen in

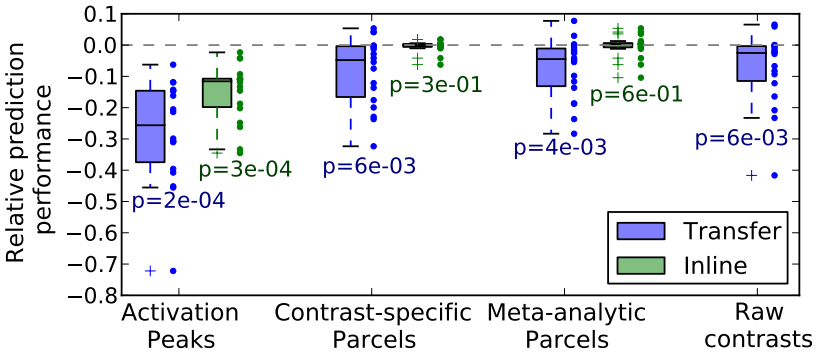


Figure 4.4: Prediction performance relative to inline prediction with raw contrasts images. The p-values indicate whether the associated methods are significantly poorer than the reference prediction method. The boxplots scatter points represent the methods' scores values within the cross-validation.

Table 4.1, the prediction scores of transfer learning as well as inline learning on this pair are acceptable although not excellent: French language and jabberwocky are difficult to separate.

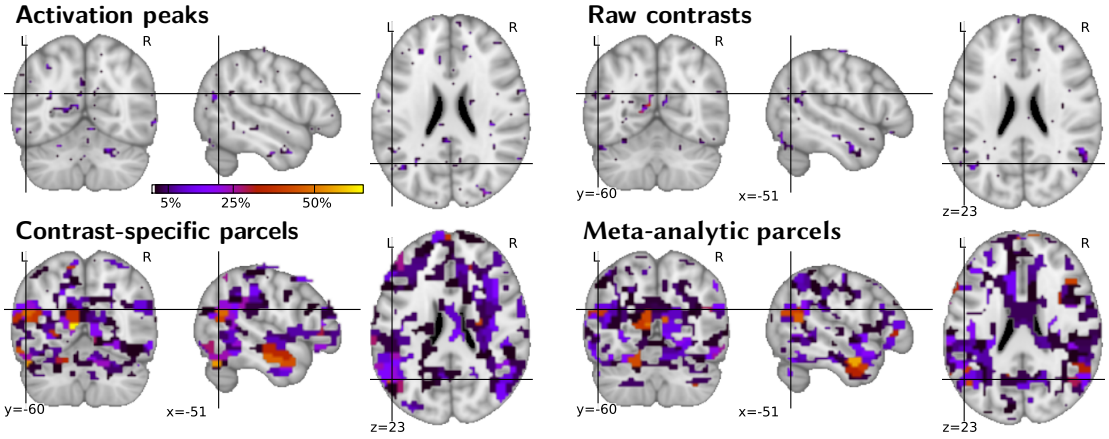


Figure 4.5: Stability scores of the randomized logistic on the Korean versus French prediction of *E2* for the different sets of features: the colormap represents the frequency at which a feature, parcel or voxel, was selected. The maps are thresholded at 1%.

Figure 4.5 gives the stability scores of the randomized logistic discriminating reading Korean from reading French for the different set of features –activation peaks, raw contrasts, parcels learned on the training contrasts or on the full database. We can see that while learning at the voxel level or at the parcel level gives similar prediction performance (Table 4.1), the stability score maps are very different. At the voxel-level, with 70 subjects ($p = 40\,000$, $n = 70$) the recovery is limited to approximately 7 voxels without randomization: the recovery conditions are violated. As a result, the randomized logistic selects only the most predictive voxels. On the parcels, contrast-specific or meta-analytic (i.e., learned on the full database), the selection frequency highlights regions of the brain that are known to be relevant for language comprehension, including the left anterior superior temporal sulcus and the part of the temporal parietal junction (Wernicke's area).

We threshold the stability selection scores of the first experiment (Korean vs French) to select candidate voxels for the target experiment (jabberwocky vs French). As we want to perform a rough screening and would rather err on the side of false detections than false rejections, we take a very low threshold $\tau = .01$. Following our analysis above, the size of the largest group of correlated features that we can detect with such a threshold is on

the order of $1/\tau \approx 100$. With 2000 parcels, this number corresponds to 5% of the brain, *i.e.* 8 000 voxels, and we can safely consider that no fMRI contrasts is composed of groups of heavily correlated features larger than this fraction.

	FWER corrected						FDR corrected					
$n^\star$	All voxels		Selection		ANOVA		All voxels		Selection		ANOVA	
10	0	(0%)	0	(0%)	0	(0%)	0	(0%)	0	(0%)	0	(0%)
20	0	(0%)	3	(0.02%)	0	(0%)	0	(0%)	4	(0.027%)	0	(0%)
40	5	(0.0084%)	33	(0.22%)	2	(0.0014%)	143	(0.97%)	1339	(9%)	201	(1.4%)

Table 4.2: Number of detections at $p < 0.05$ for difference cohort size, for transfer learning and ANOVA. The percentage of detection is indicated in parenthesis.

On the target experiment, we perform a standard group-level analysis with the voxels selected, testing for the difference between the two conditions, jabberwocky or French reading. We report results with p-values corrected for multiple comparisons at a given family-wise error rate (FWER) using Bonferroni correction, and for a given false discovery rate (FDR) using the Benjamini-Hochberg procedure. On table 4.2, we compare the number of detections and the detection rate, *i.e.* the fraction of voxels detected as significantly different, for a full brain analysis and for an analysis limited to the voxel selection. We compare our voxel selection method to a one-way ANOVA, and find that transfer learning outperforms the ANOVA for all the cohort sizes. Figure 4.6 shows the Q-Q plots on which the Benjamini-Hochberg procedure is applied. We find that voxel selection by transfer learning improves both the absolute number of detections and the detection rate for FWER and FDR correction.

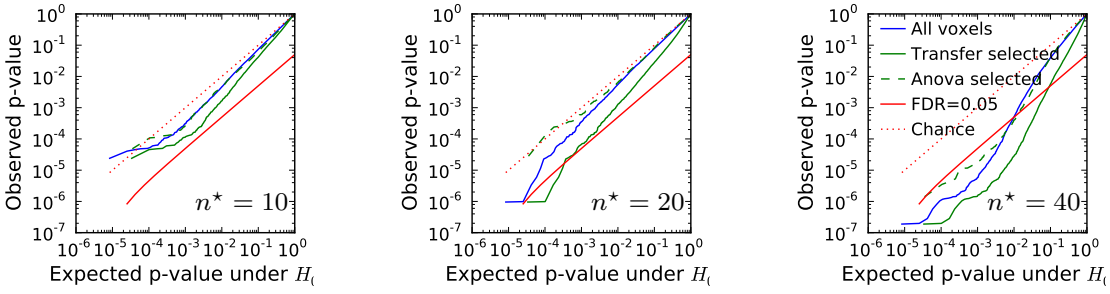


Figure 4.6: Q-Q plots for the p-values with and without voxel selection by transfer learning, as well as FDR=0.05 threshold: **left** for a cohort size $n^\star = 10$, **middle** for a cohort size $n^\star = 20$, **right** for a cohort size $n^\star = 40$.

4.1.4 Conclusion

In this approach, we propose to improve the conditioning and power of statistical analyses in imaging studies, using a large meta-analytic database.

In a *transfer learning* scheme, we train on the database sparse discriminative models that are suited to the target experiment. Not only can the predictive power of these models be useful to establish biomarkers, but also they perform feature selection that can increase the statistical power of a standard group analysis on new experiments, provided enough predictive features (voxels) can be recovered. Using brain parcellations, the discrimi-

native model filters out parcels unlikely to be relevant in the target experiment, thus defining automatically ROIs.

Using a set of 3 fMRI studies related to language, we confirm experimentally that our transfer learning scheme is able to: i) perform accurate predictions on experiments acquired on a different scanner and with varying paradigm, ii) outperform the standard meta-analysis procedures based on activation peaks, iii) increase the statistical power in the target experiment by using the ROIs defined by the discriminative model.

In this work we manually select the contrast pairs since it is delicate to interpret a transfer learning score without good knowledge of the cognitive or clinical conditions under study. The weakness of this approach lies in the arbitrary threshold we use on the classifiers' predictive features to define the regions of interest. The next section describes an approach that aims to solve this problem.

4.2 On spatial selectivity and prediction across conditions with fMRI

4.2.1 Introduction

Functional neuroimaging data are currently routinely used to better understand cognitive processes. They rely heavily on previous findings to formulate hypotheses and narrow the search space to regions of interest (ROIs), most often reported as coordinates of activation peaks [23], or from coordinate databases such as BrainMap [10]. However, understanding the literature is increasingly difficult, so that there is a need for more systematic methods, which use the images themselves to characterize the functional specificity of brain regions [19]. *Transfer learning* is a method that trains a classifier to learn a discriminant model on a source task, and then generalizes on a target task without further training. It can yield insights on some brain mechanisms if the tasks share specific common effects in some brain regions [7]. The goal of this work is to investigate the power of transfer learning procedures applied to pairs of cognitive contrasts, where the discrimination ability of the classifier quantifies the information shared between brain maps, and thus characterizes at which spatial scale functional contrasts can be jointly classified. We show that in many cases, transfer learning gives poor results in terms of spatial selectivity. To address this limitation, we introduce *selection transfer*, i.e. classification of brain states on the target task following the canonical procedure [13], but using regions defined on the source task.

4.2.2 Methods

Problem setting We start from a database holding several studies, each of them containing different functional contrast images, acquired over multiple subjects. We consider two sets of tasks, the *source tasks* and the *target tasks*, each composed of pairs of contrast images. Given n contrasts pairs of k voxel each, we call $\boldsymbol{X} \in \mathbb{R}^{n,k}$ the images of the source tasks, and $\boldsymbol{y} \in \{0,1\}^n$ the label denoting the functional contrast under study. The target images and labels are defined likewise: $\boldsymbol{X}^\star \in \mathbb{R}^{n,k}$ and $\boldsymbol{y}^\star \in \{0,1\}^n$.

The source and the target share a similar functional *spatial pattern*, and we are interested in finding the common ROIs, as well as the differences, using a machine learning approach. Note that a common pitfall in neuroimaging classification-based data processing is a successful prediction cannot guarantee that the information used by the classifier is specific to the cognitive process of interest.

Regions selection Feature selection is an important step of brain activity decoding procedures. Full brain decoding approaches are efficient but require a careful methodology to recover the contribution of different brain regions in the classification. To test the involvement of a particular brain region, researchers typically use ROIs from an atlas, or derived from the literature. Another option is to use methods such as the searchlight algorithm, in order to evaluate and extract spatially relevant voxels across the whole brain [8]. We choose to use a one-way ANOVA procedure [3], that yields a selection based on the functional activations elicited by a task, rather than using purely spatial information. We consider different fractions of the brain voxels that are most correlated to the functional contrast and perform the learning procedure on these voxels. We vary the percentiles of selected voxels with a cubic scale, from roughly 150 voxels to the full brain. This way we can control the spatial specificity against the prediction performance, and attempt to find an optimal set of regions.

Transfer learning This consists in learning discriminative models on a *source* functional task (X, y) in order to capture information that should be predictive for a *target* task $(X^\star, y^\star)$. The general assumption is that if a transfer occurs, the two experiments share at least some common cognitive circuity. Here, we train a linear classifier on the source task, and we predict the labels of the target without any additional training. The features are selected with a one-way ANOVA on the source task, which makes it possible to compare region-based transfer learning with full brain transfer learning.

Selection transfer This consists in building a predictive model for the target task based on information extracted from the source task. However, here the transfer occurs on feature selection: we perform the ANOVA procedure on (X, y) to select the most relevant voxels, then we train a linear classifier on $(X^\star, y^\star)$, and predict on the same task with the voxels selected from the source. Consequently, the transfer is not a generalization of a classifier as in transfer learning, but rather an evaluation of the significance of features from a task to another. We use the same linear classifier as the one used for transfer learning.

4.2.3 Experiments and Results

FRMI dataset

We use data from two fMRI studies for this work. The first one [17] is composed of 322 subjects and was designed to assess the inter-subject variability in some language, visual, calculation, and sensorimotor tasks. The second study is similar to the first one in terms of stimuli, but the data were acquired

on 35 pairs of twin subjects. The two studies were pre-processed and analyzed with the standard fMRI analysis software SPM5. The data used for this work are a subset of the 90 different statistical images resulting from the intra-subject analyses. The raw images were acquired on a 3T SIEMENS Trio and a 3T Brucker scanner for the first study, and on a 1.5T GE Signa for the second one. Table 4.3 presents the list of contrasts pairs used for this analysis.

Contrasts Names	Selected Scale		Area under p-curve		Description
	trans.	*sel.*	*trans.*	*sel.*	
house/scramble → face/scramble	68.11	3.25	22.73	4.51	house/scramble = house vs scrambled image
face/scramble → house/scramble	0.40	2.67	16.22	2.71	face/scramble = face vs scrambled image
word/scramble → face/scramble	23.77	4.63	10.36	2.88	word/scramble = word vs scrambled image
face/scramble → word/scramble	1.36	0.79	11.15	2.29	face/scramble = face vs scrambled image
French/sound → Korean/sound	0.40	0.02	3.57	4.61	French/sound = French listening vs sound
Korean/sound → French/sound	0.27	0.00	14.59	1.21	Korean/sound = Korean listening vs sound
V comp./sent. → A comp./sent.	11.01	0.00	2.62	1.76	V comp./sent. = computation vs reading
A comp./sent. → V comp./sent.	0.01	6.36	4.75	3.10	A comp./sent. = computation vs listening
V motor/sent. → A motor/sent.	0.10	0.00	11.84	1.85	V motor/sent. = button press vs reading
A motor/sent. → V motor/sent.	7.37	0.00	4.45	2.11	A motor/sent. = button press vs listening

Table 4.3: Source and target tasks: **Selected scales** and **area under the p-values curve** for both transfer learning and selection transfer. trans.= transfer learning; sel.= selection transfer; V= visual stimuli; A= auditory stimuli.

Experimental results for transfer learning

We are interested in *transfer learning*: we learn a discriminative model on the source task with a univariate feature selection, and predict the labels on the target task.

The analysis presents two phases: we first train a linear classifier on a source task, and then re-use the discriminative model on the target task to perform the *transfer learning*; this is repeated on 6 different sub-samples of the source task to estimate the uncertainty on transfer accuracy. We use two kinds of linear classifiers: a SVC (Support Vector Classifier) and a Logistic Regression with ℓ_2 penalization. The penalization is set by nested 6-fold cross-validation for each classifier. We find that the two methods yield very close results, and thus report only results using the SVC classifier. We also train and then test the classifier on the target task and call this procedure *inline learning*. In Figure 4.7, we show the performance τ_p^t of transfer learning, relative to inline learning τ_p^i, varying the percentile p of features selected in a cubic scale. In general, for any given p, τ_p^i can remain significantly higher than τ_p^t. For this reason, we use a heuristic to select the scale parameter (see also Figure 4.7): the scale that yields the minimal $\tau_p^i - \tau_p^t$ difference. We consider that at this scale, the maps associated with the two tasks share a maximal amount of common information.

However, the voxels selected with this method are either too few to give an accurate prediction, or too many to yield identifiable regions. The transfers do not behave the same way on both directions: in general, one direction is more sensitive but less specific, and the other direction shows the opposite behaviour. This comes from tasks-related foci being more spatially focused for some contrasts. Because of this lack of specificity, we do not find contained regions that overlap with the Fusiform Face Area (FFA) [6], the Parahippocampal Place Area (PPA) [5] or the Visual Word Form

Area (VWFA) [2], regions respectively involved in face recognition, object visual processing, and reading.

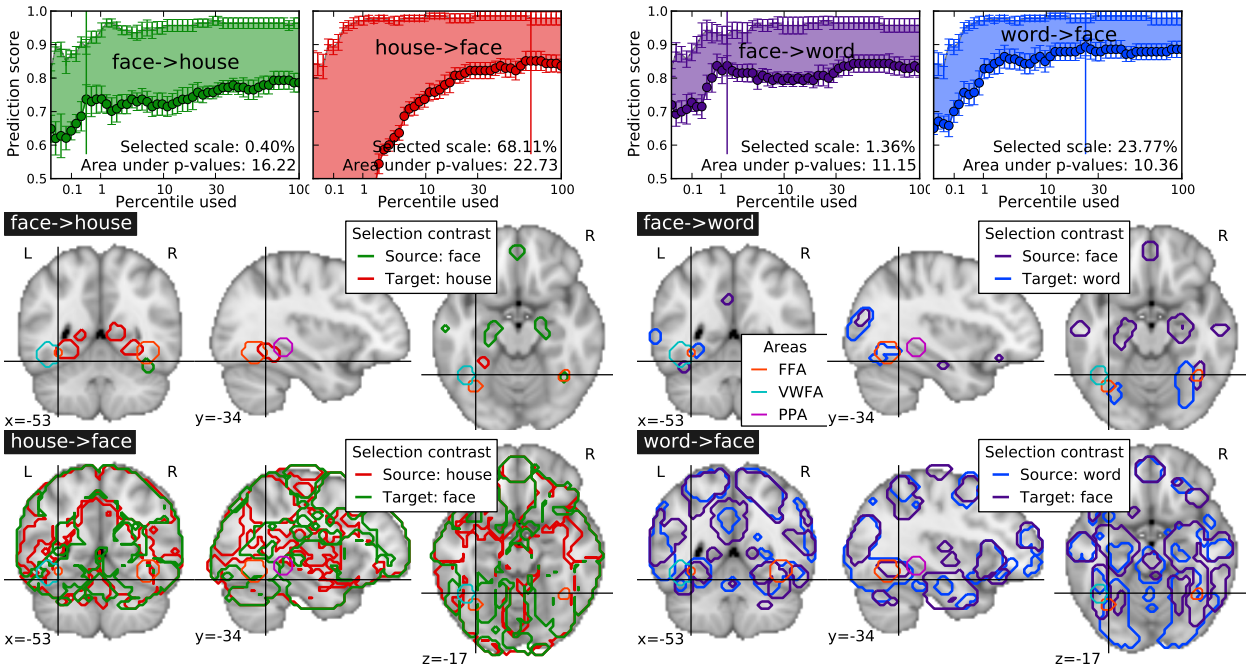


Figure 4.7: Example of results using the **Transfer learning** approach, in four different transfer settings: we can see that the area between the inner transfer prediction accuracy curves are large, and that the prediction rates do not converge. The *optimal scale*, defined as the minimum of the difference between the curves, often corresponds to a rather broad, non-specific brain map.

Experimental results for selection transfer

We are interested in *selection transfer*: we do not perform transfer learning, instead, we use the univariate feature selection performed on the source task, to learn a discriminative model and predict the labels in the target task.

We use the same machine learning tools as the transfer learning: we train and test a linear classifier with a 6-fold cross validation test on the target task. For this method the SVC and the Logistic Regression with ℓ_2 penalization also give very close results. As with *transfer learning*, we also perform an inline learning on the target task, with features selected on the same images.

On Figure 4.8, we show the performance τ_p^s of selection transfer against inline learning τ_p^i, and how the performance varies with the percentile p of the brain recruited for the learning process. In comparison to transfer learning, two things happen: *i)* the selection transfer is more symmetric, *ii)* τ_p^i is not significantly higher than τ_p^s for every p. We can therefore use a t-test to define the *selected scale* (Figure 4.8) as the first one with non significant difference between the curves. This enables us to control the amount of information to include in the prediction problem, and have both a good performance and an improved specificity of the regions selected for the two tasks. In practical terms, the *selected scale* makes it possible to identify the smallest fraction of the brain that yields overlapping regions in the two tasks, and consequently an accurate prediction. Although the selected regions have no guarantee of optimality, they are specific enough to overlap with the FFA, the PPA and the VWFA. We can also use the area under the

p-values curve from the t-test as a measure of similarity between the tasks. While it is not possible to interpret this measure absolutely, we can use it to compare one task versus others. For the example on Figure 4.8, we can see that the area between *face* and *word* is smaller than between *face* and *house*. This indicates that the face task is closer to the word task than the house task, which is consistent with previous findings [4].

Limitations *Selection transfer* captures voxels that generalize well in terms of prediction from one task to another. However, a classifier may require very few voxels to perform well, in which case this method misses some regions involved in the cognitive process of interest. This effect is represented by the values in Table 4.3, where *selection transfer* requires only a small p fraction of the brain to obtain a τ_p^s, which is not significantly lower than τ_p^i (e.g., V comp./sent → A motor/sent.). In order to retrieve optimal regions when this is the case, a standard analysis, based either on contrast addition or conjunction [14], would be sensitive enough to detect the common active regions for both tasks.

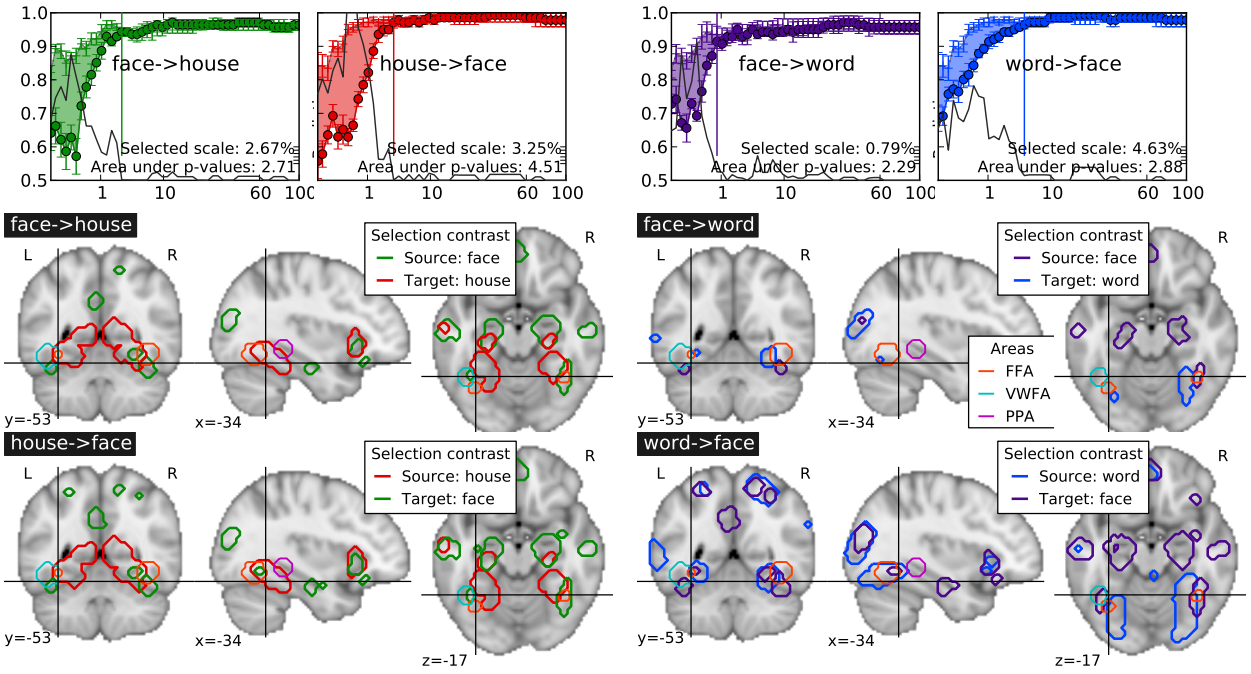


Figure 4.8: Example of result using the **Selection transfer** approach: The two prediction curves do converge, so that the difference becomes non-significant as soon as a relatively small fraction of the voxels are included: the spatial scale is defined here as the point where the curves can no longer be distinguished. It corresponds to more symmetric and meaningful brain maps than those obtained with transfer learning.

4.2.4 Conclusion

In this contribution, we investigate the ability of *transfer learning* and *selection transfer* to characterize the spatial scale at which functional contrasts can be jointly classified. The objective is to find a systematic procedure to extract ROIs that define common information between two functional tasks, instead of relying on activation coordinates from the literature. We show that transfer learning does not provide control of the regions' size it uses to classify the tasks. Instead we use a selection transfer procedure that seems to better characterize which fraction of the brain yields discriminant information. Our results suggest that transfer learning requires to be used in a

carefully designed study, as it is difficult to control the spatial selectivity of this method. Another interesting result is that selection transfer is not symmetric (i.e., source and target tasks are not inversible), as opposed to contrast conjunction.

4.3 Conlusion

The methods presented in this chapter aim to use pairs of tasks to define common regions of interest. The first contribution examines the ability of one task to better condition testing for a second similar task. The second contribution both automates the threshold setting from the first method, and proposes a multivariate alternative to conjunction, which remains a valid way to leverage multiple tasks for inference. In hindsight, it would have been more straightforward to estimate a null-hypothesis by permutation testing, and define regions by assessing statistical significance. It would however be computationally expensive, and would not address the main limitation of these contributions, which is their ability to scale: we do not have any means to automate the selection of task pairs to compare, as a classification score alone is not a good metric to assess task similarity.

Bibliography

[1] F. Bach. Self-concordant analysis for logistic regression. *Electronic Journal of Statistics*, 4:384–414, 2010.

[2] Laurent Cohen and Stanislas Dehaene. Specialization within the ventral stream: the case for the visual word form area. *NeuroImage*, 22:466, 2004.

[3] D. D. Cox and R. L. Savoy. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19:261, 2003.

[4] Stanislas Dehaene, Felipe Pegado, Lucia W Braga, et al. How learning to read changes the cortical networks for vision and language. *Science*, 330:1359, 2010.

[5] R Epstein and N Kanwisher. A cortical representation of the local visual environment. *Nature*, 392:598, 1998.

[6] N. Kanwisher, J. McDermott, and M.M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.*, 17:4302, 1997.

[7] André Knops, Bertrand Thirion, Edward M Hubbard, Vincent Michel, and Stanislas Dehaene. Recruitment of an area involved in eye movements during mental arithmetic. *Science*, 324:1583, 2009.

[8] Nikolaus Kriegeskorte, Rainer Goebel, and Peter Bandettini. Information-based functional brain mapping. *Proc Ntl Acad Sci*, 103(10):3863, 2006.

[9] Angela R Laird, P. Mickle Fox, Cathy J Price, David C Glahn, Angela M Uecker, Jack L Lancaster, Peter E Turkeltaub, Peter Kochunov, and Peter T Fox. Ale meta-analysis: controlling the false discovery rate and performing statistical contrasts. *HBM*, 25(1):155, 2005. doi: 10.1002/hbm.20136.

[10] Angela R Laird, Jack L Lancaster, and Peter T Fox. Brainmap: the social evolution of a human brain mapping database. *Neuroinformatics*, 3(1):65–78, 2005.

[11] N. Meinshausen and P. Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

[12] V. Michel, A. Gramfort, G. Varoquaux, E. Eger, C. Keribin, and B. Thirion. A supervised clustering approach for fMRI-based inference of brain states. *Pattern Recognition*, 2011.

[13] Janaina Mourão-Miranda, Arun L.W. Bokde, Christine Born, Harald Hampel, and Martin Stetter. Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. *NeuroImage*, 28:980, 2005.

[14] T. Nichols, M. Brett, J. Andersson, T. Wager, and J.B. Poline. Valid conjunction inference with the minimum statistic. *Neuroimage*, 25:653, 2005.

[15] Christophe Pallier, Anne-Dominique Devauchelle, and Stanislas Dehaene. Cortical representation of the constituent structure of sentences. *PNAS*, 108(6):2522–2527, 2011. doi: 10.1073/pnas.1018711108.

[16] S.J. Pan and Q. Yang. A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions on*, 22 (10):1345–1359, 2010.

[17] Philippe Pinel, Bertrand Thirion, Sébastien Meriaux, Antoinette Jobert, Julien Serres, Denis Le Bihan, Jean-Baptiste Poline, and Stanislas Dehaene. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci*, 8:91, 2007. doi: 10.1186/1471-2202-8-91.

[18] Gholamreza Salimi-Khorshidi, Stephen M Smith, John R Keltner, Tor D Wager, and Thomas E Nichols. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *Neuroimage*, 45(3):810–823, 2009. doi: 10.1016/j.neuroimage.2008.12.039.

[19] A.J. Sutton, K.R. Abrams, DR Jones, TA Sheldon, and F. Song. Methods for meta-analysis in medical research. *West Sussex, UK: Chichester, England*, 2000.

[20] Bertrand Thirion, Philippe Pinel, Sébastien Mériaux, Alexis Roche, Stanislas Dehaene, and Jean-Baptiste Poline. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage*, 35(1): 105–120, 2007. doi: 10.1016/j.neuroimage.2006.11.054.

[21] Peter E Turkeltaub, Guinevere F Eden, Karen M Jones, and Thomas A Zeffiro. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage*, 16(3 Pt 1):765–780, 2002.

[22] Tor D Wager, Martin A Lindquist, Thomas E Nichols, Hedy Kober, and Jared X Van Snellenberg. Evaluating the consistency and specificity of neuroimaging data using meta-analysis. *Neuroimage*, 45(1 Suppl):S210–S221, 2009. doi: 10.1016/j.neuroimage.2008.10.061.

[23] Tal Yarkoni, Russell A Poldrack, Thomas E Nichols, David C Van Essen, and Tor D Wager. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods*, 8(8):665–670, 2011. doi: 10.1038/nmeth.1635.

5 Learning functional networks

This chapter takes a step back from multi-study analyses. Experiments from the previous chapter indicate that classifier generalization is a useful method to compare functional tasks in a controlled setting, but it does not provide a way to scale and therefore benefit from data accumulation. We previously focused on regions of interest but more modern approaches view the brain functioning within a set of distributed networks [23]. Such functional networks have been successfully exposed in resting state fMRI [17, 2], as well as task fMRI data [3, 11]. These works primarily rely on unsupervised decomposition methods such as Independent component analysis (ICA) or Principal Component Analysis (PCA), that brought the resting state networks to light [5]. The main pitfall remains the **labeling of the functional networks**, that is done manually by associating brain maps with known brain functions.

The contribution described in this chapter proposes an alternative way to expose functional networks on task fMRI, and associates them with what we call **functional profiles**. We rely on another unsupervised approach, that learns jointly a dictionary of functional profiles and a set of spatial maps [10]. fMRI studies outline mental processes by combining experimental conditions. The functional profiles reflect the response magnitude of the experimental conditions for each network, and permit to define a functional signature. In particular, this contribution investigates three dictionary learning approaches, that encode respectively spatial matching between subjects, functional matching, and random effect modeling of the intra and inter-subject variance.

The contribution developed in this chapter have been published in:

- G. Varoquaux, Y. Schwartz, P. Pinel, and B. Thirion, *Cohort-level brain mapping: learning cognitive atoms to single out specialized regions*, IPMI, 2012, pages 438–449.

Contents

5.1 Introduction

Using fMRI, the systematic study of which areas of the brain are recruited during various experiments has led to accumulation of activation maps related to specific tasks or cognitive concepts in an ever growing literature. Mapping a given population requires careful crafting of a set of tasks that are contrasted to reveal networks. These networks form a natural representation of brain function and are of particular interest to study its variability in a population, for instance to correlate it to pathologies or genetic information. However, each subject can only perform a small number of tasks in a scanner; particularly so for disabled subjects. As a result, in a given study the number of networks that are identified by standard task-activation mapping is small and limited by the number of contrasts of the study. On the other hand, it is not uncommon to scan a large number of subjects. Indeed, clinical studies must often resort to larger sample sizes due to the intrinsic variability of pathologies. Massive cohorts can be acquired, *e.g.* to learn diagnosis markers for Alzheimer's disease [13], or in neuroimaging-genetics.

In large cohorts, a small set of contrasts reveals effects throughout the whole brain [20]. This observation suggests that more information can be extracted at the cohort level. In this paper, we address precisely this challenge by decomposing brain activity at the group level to assign a specific cognitive function to each voxel. For this purpose, inter-subject variability is a blessing as functional variability reveals *functional degeneracy*, *i.e.* that different networks sustain the same cognitive function across individuals [12]. However, this variability is also a curse when it arises from spatial realignment error.

Compressed spatial representations were put forward for group studies by Thirion *et al.* [19] using clustering of the activation maps. This early work did not address the functional specificity of the clusters. Conversely, Lashkari *et al.* [9] discard spatial information and focus on extracting common functional profiles across subjects, removing the need for spatial normalization. Following this idea of functional correspondence across subjects, although not leading to the definition of regions, Sabuncu *et al.* [15] use this correspondence for inter-subject alignment. Linear models such as independent component analysis (ICA) have been used to extract modes of brain function across subjects [3] before clustering approaches. Laird *et al.* [8] have recently shown that the modes that it extracts from task-activation

data capture meaningful structure in the space of cognitive processes. Beyond ICA, Varoquaux *et al.* [22] use dictionary learning to segment a functional parcellation from resting-state. Very interesting preliminary work by Chen *et al.* [4] integrates spatial normalization with dictionary learning to estimate jointly an inter-subject warping and functional regions.

This chapter combines ideas from this prior art in a new inter-subject model with an associated computationally-scalable estimation algorithm. Our contributions are *i)* a joint model of the position and functional tuning of brain networks, *ii)* explicit separation of the variance into intra-subject and inter-subject components, *iii)* a fast and scalable algorithm that can impose this particular variance structure. We show with simple simulations that controlling inter-subject variance is crucial, as unsupervised learning approaches such as dictionary learning or clustering will fit this variance and extract modes reflecting inter-subject variability. The paper is organized as followed. We start by giving a multi-subject model combining random effects (RFX) with functional segregation hypothesis. In section 5.3, we introduce an on-line and computationally-efficient algorithm to estimate this model. In section 5.4, we present a simulation study, and in section 5.5 results on an fMRI dataset comprising 150 subjects.

5.2 A multi-subject sparse-coding model of brain response

Sparse coding brain response Our model is based on two basic neuroscience principles: *i) functional segregation* which states that brain territories are formed of elementary, functionally-specific units [21] and *ii) functional degeneracy* which states that a particular function may recruit different networks across subjects [12]. We combine these principles at the subject and group level to learn the correct basis to describe the macroscopic level of brain organization.

Experimental stimuli and contrasts do not correspond simply to elementary cognitive processes. For instance to isolate brain regions involved in a calculation tasks, instructions to perform arithmetics will be given to a subject, however these instructions are given via a modality: auditory or visual, and will induce a word-comprehension task in addition to the calculation. Investigators use *contrast maps* to cancel out secondary effects and focus on *word – calculations*, but these contrasts can carry also some auditory, visual, or language effects as the stimuli content in the different tasks are not perfectly matched.

A typical fMRI experiment thus yields a set of task-specific contrast maps: for each subject s, $\boldsymbol{X}^s \in \mathbb{R}^{t \times n}$, where t is the number of tasks and n the number of voxels. Based on the principles of functional specialization, we stipulate that the tasks used are formed of elementary cognitive processes associated with a set of corresponding sparse neural substrates: there exist combinations of tasks $\boldsymbol{D} = \{\boldsymbol{d}_j\}$ such that each $\boldsymbol{d}_j$ is expressed on a small number of brain regions:

$$\boldsymbol{X}^s = \boldsymbol{D}\boldsymbol{A}^{s\mathsf{T}}, \qquad \text{where } \boldsymbol{A}^s \text{ is sparse.} \tag{5.1}$$

We are interested in learning a dictionary of k functional profiles $\boldsymbol{D} \in \mathbb{R}^{t \times k}$ and the associated sparse spatial code $\boldsymbol{A}^s \in \mathbb{R}^{n \times k}$, that we call *functional*

networks. The number of atomic cognitive functions recruited by the tasks explored in an fMRI experiment is most likely much larger than the number of experimental conditions t. Drawing from a large number of subjects can help to estimate more functional profiles, as subjects will resort to different *cognitive strategies*, engaging differently atomic cognitive functions. To give a clichéd image, right-handed and left-handed subjects could rely on different visuo-spatial representations to perform a hand motion task. In practice, variability in cognitive strategy is often very subtle and can be related to variability in attention, engagement to the task, background processes, rather than high-level strategies [12]. Modeling this inter-subject variability should improve the quantification of population-level estimates and enable the separation of atoms of brain function.

Multi-subject modeling We introduce subject-specific expressions of the functional profiles:

$$\boldsymbol{F}^s = (\boldsymbol{I} + \Delta^s)\boldsymbol{D}, \qquad \text{where } \Delta^s \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_t), \quad \Delta^s \in \mathbb{R}^{t\times t}, \quad \boldsymbol{F}^s \in \mathbb{R}^{t\times k} \tag{5.2}$$

An approach commonly used when dealing with such unsupervised learning problem on multi-subject fMRI data is to concatenate the data spatially [3, 19], learning an augmented dictionary,

$$\overline{\boldsymbol{F}} \;=\; [\boldsymbol{F}^{1\mathsf{T}} \dots \boldsymbol{F}^{s\mathsf{T}}]^\mathsf{T} \;=\; [(\boldsymbol{I} + \Delta^1)^\mathsf{T}, \dots (\boldsymbol{I} + \Delta^s)^\mathsf{T}]^\mathsf{T}\boldsymbol{D} \;\in \mathbb{R}^{st\times k}. \tag{5.3}$$

The multi-subject model resulting from (5.1) and (5.2) can then be written as a standard dictionary-learning problem: $\overline{\boldsymbol{X}} = \overline{\boldsymbol{F}}\boldsymbol{A}^\mathsf{T}$, with $\overline{\boldsymbol{X}} \in \mathbb{R}^{st\times n}$ the spatial concatenation of the data and $\boldsymbol{A}$ functional networks independent of the subject. By learning a dictionary spanning multiple datasets, it can estimate inter-subject loadings that reveal the different cognitive strategies, drawing from the *spatial correspondence* of the coding of the information. However, estimating high-dimensional dictionaries has two major drawbacks: *i)* it is more challenging from the statistical standpoint because the residuals implicit in eq. 5.3 are non white and *ii)* this approach is fragile to errors in inter-subject correspondence.

To remove the need for spatial matching, Lashkari *et al.* [9] cluster the activity profiles, grouping voxels that respond similarly to the tasks across subjects. This *functional correspondence* hypothesis leads to a functional concatenation of the data: $\underline{\boldsymbol{X}} = [\boldsymbol{X}^1, \dots \boldsymbol{X}^{s\mathsf{T}}]^\mathsf{T} \in \mathbb{R}^{t\times sn}$. The multi-subject model is then written $\underline{\boldsymbol{X}} = \boldsymbol{D}\underline{\boldsymbol{A}}^\mathsf{T}$ with $\underline{\boldsymbol{A}} = [(\boldsymbol{I}_k + \Delta^1)\boldsymbol{A}^1, \dots (\boldsymbol{I}_k + \Delta^s)\boldsymbol{A}^{s\mathsf{T}}]^\mathsf{T} \in \mathbb{R}^{k\times sn}$, which amounts to learning a dictionary common to all subjects and different spatial maps.

Modeling Random effects Both spatial and functional concatenation approaches lead to a simple formulation in terms of learning a dictionary of functional profiles and spatial code. However a naive resolution of these dictionary learning problems neglects that both spatial code and functional profiles share information across subjects. In functional neuroimaging data analysis, the standard way to model both common effects and variability across datasets relies on hierarchical linear models, often mixed- or random-effects (RFX) models that assume that the effect has two components of

variance: inter-subject and intra-subject [24]. We can adapt this model to enhance the spatial correspondence approach by constraining the ratio of the intra- and inter-subject variance of the functional profiles in the augmented dictionary $\overline{\boldsymbol{F}}$. For this purpose, we introduce a *common effect matrix* made of s $k \times k$ identity matrices concatenated: $\boldsymbol{C} = \frac{1}{s}[\boldsymbol{I}_k, \dots \boldsymbol{I}_k]^\mathsf{T} \in \mathbb{R}^{k \times sk}$ and the *differential effects matrix* $\boldsymbol{C}_\perp \in \mathbb{R}^{(s-1)k \times sk}$, which is an orthogonal completion of $\boldsymbol{C}$. To impose an RFX structure on the dictionary, we present in section 5.3 an algorithm controlling $\|\overline{\boldsymbol{f}}_i \boldsymbol{C}\|_2^2 / \|\overline{\boldsymbol{f}}_i \boldsymbol{C}_\perp\|_2^2$, where $i \in [1, t]$ is the index of a dictionary element.

Proposition 1. *$\boldsymbol{C}$ and $\boldsymbol{C}_\perp$ isolate* i) *group-level profiles:* $\mathbb{E}[\overline{\boldsymbol{f}}_i \boldsymbol{C}] = \boldsymbol{d}_i$,
ii) *intra-subject variance:* $\mathbb{E}[\|\overline{\boldsymbol{f}}_i \boldsymbol{C}\|_2^2] = \left(1 + \frac{\sigma^2}{s}\right)\|\boldsymbol{d}_i\|_2^2 \approx \|\boldsymbol{d}_i\|_2^2$,
iii) *inter-subject variance:* $\mathbb{E}[\|\overline{\boldsymbol{f}}_i \boldsymbol{C}_\perp\|_2^2] = \left(\sigma^2 - \frac{\sigma^2}{s}\right)\|\boldsymbol{d}_i\|_2^2 \approx \sigma^2 \|\boldsymbol{d}_i\|_2^2$.

The first and the second equalities stem from Eq. (5.3), while the last one follows from the fact that $\|\overline{\boldsymbol{f}}_j\|_2^2 = \|\overline{\boldsymbol{f}}_j \boldsymbol{C}\|_2^2 + \|\overline{\boldsymbol{f}}_j \boldsymbol{C}_\perp\|_2^2$, as $[\boldsymbol{C}^\mathsf{T}, \boldsymbol{C}_\perp^\mathsf{T}]$ forms an ortho normal basis of $\mathbb{R}^{sk}$.

5.3 Efficient learning of RFX-structured dictionaries

State-of-the-art dictionary learning algorithm A general approach to learn dictionaries for sparse coding is to optimize the dictionary so that is leads to a sparse regression on train data, using an ℓ_1 penalty on the code [10]:

$$\hat{\boldsymbol{D}} = \underset{\boldsymbol{A}, \boldsymbol{D},\, \boldsymbol{D} \in C}{\operatorname{argmin}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{A}^\mathsf{T}\|_2^2 - \lambda \|\boldsymbol{A}\|_1, \tag{5.4}$$

where $\boldsymbol{X}, \boldsymbol{D}, \boldsymbol{A}$ should be replaced by $\overline{\boldsymbol{X}}, \overline{\boldsymbol{F}}, \boldsymbol{A}$ or $\underline{\boldsymbol{X}}, \boldsymbol{D}, \underline{\boldsymbol{A}}$ depending on the choice of correspondence. Note that the dictionary $\boldsymbol{D}$ is constrained to a convex set C, typically by bounding the ℓ_2 norm of its atoms: $\|\boldsymbol{d}_i\|_2 \leq 1$. This constraint is technical, as without it the penalty on $\boldsymbol{A}$ could be made arbitrarily small by scaling up $\boldsymbol{D}$ and down $\boldsymbol{A}$ and thus keeping the data-fit term constant. Let us rewrite the optimization problem:

$$\hat{\boldsymbol{D}} = \underset{\boldsymbol{D},\, \boldsymbol{D} \in C}{\operatorname{argmin}} \sum_v \min_{\boldsymbol{a}_v} \left(\left\| \boldsymbol{x}_v - \boldsymbol{D}\boldsymbol{a}_v^\mathsf{T} \right\|_2^2 + \lambda \|\boldsymbol{a}_v\|_1 \right). \tag{5.5}$$

This new expression highlights that, when learning the dictionary, the objective function is the sum over a large number of different realizations of the same problem, here sparse coding a simple voxel activation profile $\boldsymbol{x}_v$. The optimization problem can thus be tackled using stochastic gradient descent with on-line or mini-batch strategies [10]: small numbers of voxels randomly drawn from the data are successively considered and a corresponding sparse code $\boldsymbol{a}_v$ is learned by solving a Lasso-type problem. The dictionary can then be updated to minimize the data-fit error given the code. The algorithm iterates over small batches of voxels (hundreds) to incrementally improve the dictionary. When the number of voxels is large, such an approach can be orders of magnitude faster than the alternate optimization strategies used by [22, 4], because these require solving brain-wide sparse regression for each update of the dictionary.

Szabo *et al.* [18] extend this approach to structured dictionaries by replacing the ℓ_1 norm on $\boldsymbol{\alpha}_v$ with a structure-inducing norm, such as the ℓ_{21}

norm used in the group lasso. However, the corresponding algorithms to learn the sparse code $\boldsymbol{a}_v$ are much more costly as they rely in general on optimizing augmented problems over auxiliary variables [18]. On the opposite, efficient algorithms to solve the ℓ_1 problem benefit from the sparsity of the solution and can be much less costly than a least-square estimate for very sparse problems [6].

Imposing RFX-structured dictionaries We introduce a simple modification to the on-line algorithm [10] to impose an RFX structure on the dictionary. Our approach is based on spatial correspondence to learn an augmented dictionary $\overline{\boldsymbol{F}}$ and sets different intra and inter-subject variance using proposition 1: controlling the ratio of the norm of $\overline{\boldsymbol{F}}\boldsymbol{C}$ and $\overline{\boldsymbol{F}}\boldsymbol{C}_\perp$. For this purpose, we use a careful choice of constraint set $\boldsymbol{C}$ on the dictionary; namely, we impose on each atom

$$\Omega(\overline{\boldsymbol{f}}_i) \leq 1, \qquad \text{with } \Omega(\overline{\boldsymbol{f}}_i) = \max(\|\overline{\boldsymbol{f}}_i\boldsymbol{C}\|_2^2,\ \mu\|\overline{\boldsymbol{f}}_i\boldsymbol{C}_\perp\|_2^2), \tag{5.6}$$

where μ controls the ratio of intra to inter subject variance. Because of the penalty on $\boldsymbol{A}$, it is highly likely that the constraint will be saturated. This constraint is an ℓ_∞ norm, which tends to enforce equality when saturated[1]: $\|\overline{\boldsymbol{f}}_i\boldsymbol{C}\|_2^2 = \mu\|\overline{\boldsymbol{f}}_i\boldsymbol{C}_\perp\|_2^2$.

In the on-line dictionary learning algorithm, this constraint is enforced by an Euclidean projection (see algorithm 2 of [10]): at each iteration

$$\boldsymbol{d}_{n+1} \leftarrow \operatorname*{argmin}_{\boldsymbol{d}} \|\boldsymbol{d}_n - \boldsymbol{d}\|_2^2 \quad \text{subject to } \Omega(\boldsymbol{d}) \leq 1. \tag{5.7}$$

The *max* operator in Ω imposes that $\|\overline{\boldsymbol{f}}_i\boldsymbol{C}\|_2^2 \leq 1$ and $\|\overline{\boldsymbol{f}}_i\boldsymbol{C}_\perp\|_2^2 \leq \frac{1}{\mu}$. As $\boldsymbol{C}$ and $\boldsymbol{C}_\perp$ span orthogonal subspaces, the Euclidean distance decomposes in two independent optimization problems on those subspaces: the projection on a ball of radius 1 (resp. $\frac{1}{\mu}$), $\boldsymbol{c}_{n+1} \leftarrow \boldsymbol{c}_n / \|\boldsymbol{c}_n\|_2$, where $\boldsymbol{c}$ is the restriction of $\boldsymbol{d}$ to the subspace spanned by $\boldsymbol{C}$ (resp. $\boldsymbol{C}_\perp$). In practice, to implement this projection, we apply the dictionary-update algorithm after rotating the dictionary and the code to express them in the basis of $\mathbb{R}^{sk}$ formed by $[\boldsymbol{C}^\mathsf{T}, \boldsymbol{C}_\perp^\mathsf{T}]$, and for the sparse-coding step, we rotate back the dictionary to the basis that leads to sparse codes. With this strategy, the Euclidean projection Eq. (5.7) has the same computational cost with norm Ω than with the standard ℓ_2 norm proposed in [10]. As the computational cost of the dictionary update step is already quadratic in the length of the atoms, this strategy to impose an RFX structure on the dictionary does not change the overall algorithmic complexity of dictionary learning, neither asymptotically nor for small dictionaries.

Parameter choice and initialization Our algorithm has two parameters: λ, that controls the sparsity of the spatial maps, and μ that controls the ratio of intra-subject to inter-subject variance. We set that ratio to 10. Typically in fMRI study, inter-subject variance is 4 to 9 times larger than intra-subject variance [24], thus we are over-penalizing. However, in statistics, over-penalization is considered as preferable to under-penalization, as the former leads to bias, here to a common effect, while the later can easily lead to an explosion of variance. With regards to λ, the natural scaling

[1] Indeed, combined with an ℓ_2 loss, an ℓ_∞ constraint tends to saturate at its *kinks*, enforcing equality between variables as an ℓ_1 constraint enforces sparsity.

factor is $\lambda \propto \frac{1}{\sqrt{p}}\varepsilon$ where p is the size of the atoms, and ε^2 the variance of the residuals [1]. We assume that $\varepsilon^2 \propto \operatorname{std} \boldsymbol{X}$ and use the simple choice $\lambda \propto \frac{1}{\sqrt{p}}\operatorname{std} \boldsymbol{X}$. Similar scalings are suggested in [10]. They lead to having a number of non-zero constant on average in the code $\boldsymbol{A}$. In other words, each voxel is coded on the same number of maps, independently of the size of the problem (number of maps extracted, number of contrasts).

The dictionary learning problem is not convex. The starting point is important because a good choice can significantly speed up the convergence, and also determine the final results. We use spatially-constrained clustering on spatially-concatenated data [19] to learn an initial parcellation and associated dictionary.

5.4 Results on simulated data

Synthetic data generation We generate a simple and well-understood synthetic dataset to illustrate how the different approaches work, as well as the impact of spatial variability. We study the scenario in which two observed contrasts are generated from three functional networks, each one of them made of a single blob (Fig. 5.1, top left). Group-level loadings are generated from a uniform $[0, 1]$ distribution, and for each subject one cognitive strategy out of two, corresponding to a variation in 20% of the weights, is affected randomly. Finally, Gaussian-distributed noise is added with a variance of 0.1. We generate images of size 50×50 for 32 subjects. Optionally, we add spatial variability across subjects with Gaussian noise of 3 pixel standard deviation on the positions of the blobs.

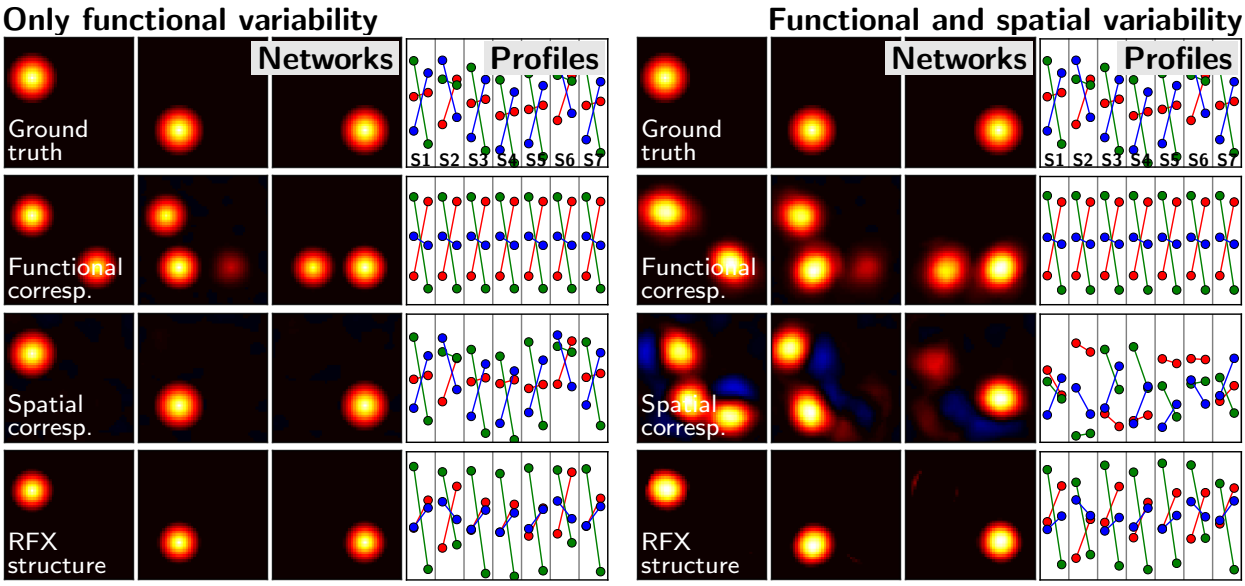


Figure 5.1: Simulations: functional networks and subject-level profiles as estimated by different dictionary learning strategies – right column: with only functional variability – left column: with spatial variability. On the ground-truth profile plot the second cognitive strategy can be seen from the red loadings in the second and sixth subjects.

Results Without spatial variability, spatial correspondence and RFX structure are very successful at singling out the blobs, however the functional correspondence strategy is less so (see Fig. 5.1). This is not surprising, as in the functional correspondence case, the dictionary learning task amounts

to separating out 3 vectors (functional profiles) in a 2-dimensional space, which corresponds to an under-determined source separation problem. The under-determined problem is much harder than the over-determined problem, as in the spatial correspondence approach. Indeed, learning an augmented dictionary across subjects can benefit from inter-subject functional variability to tease out networks. However, in the presence of spatial variability, the simple spatial correspondence fits this variability and the estimated maps exhibit *adjustment* modes, combining different networks with negative regions that correspond to network spatial derivatives. Indeed, the loadings show little consistency across subjects, as the learned spatial maps are combined to compensate for spatial fluctuations. The RFX structure prevents such a combination to happen via a shrinkage to common factors. As a result the spatial maps are more faithful to the true networks. Note that the inter-subject profiles are overly shrunk. This an expected consequence of strong regularization: suppressing the variance comes to the cost of a bias. However this bias is not detrimental to the mean profile or the spatial maps.

5.5 Learning a cognitive brain atlas from fMRI

Functional localizer dataset We use a functional localizer that targets a wide spectrum of cognitive processes, namely visual, auditory and sensorimotor processes, as well as reading, language comprehension and mental calculation. This protocol [14] lasts only 5 minutes, in order to be performed routinely on top of other protocols. We use 151 subjects that were acquired on the same 3T SIEMENS Trio scanner. 6 contrast maps best represent the brain activity for the cognitive processes recruited in this protocol. The contrast maps are both a combination of several conditions (e.g., sentence reading, calculation), and a difference of those conditions (e.g., right click versus left click) to draw out the effect of interest. For instance, the map "words - calculation" aims to isolate the effect of calculation by canceling out the modality of the stimulus (auditory or visual), and the residual effect of the comprehension of the stimulus (reading or listening). The effect of words is then encoded by negative loadings.

Networks and profiles extracted Fig. 5.2 shows some functional networks and profiles extracted using $k = 50$. The profiles are represented by their loadings on the contrasts of the original experiment, that oppose one type of brain function to another. Some networks extracted correspond across methods: for instance the network corresponding to a left click (a1, b1 and c1), for which the spatial map highlights the hand area of the motor cortex and the functional profiles are concentrated on the motor and left contrasts. As finger movement gives very strong activations, this network is reliable across subjects: standard errors on contrast loadings are small and the inter-subject functional profiles (Fig. 5.3) are similar across subjects even without enforcing structure. Note that a similar right-click network is also extracted (not shown). Extracting such a network is no surprise, as it maps well to a task performed in the study. More interestingly, networks corresponding to higher-level cognition are also extracted, *e.g.* the language

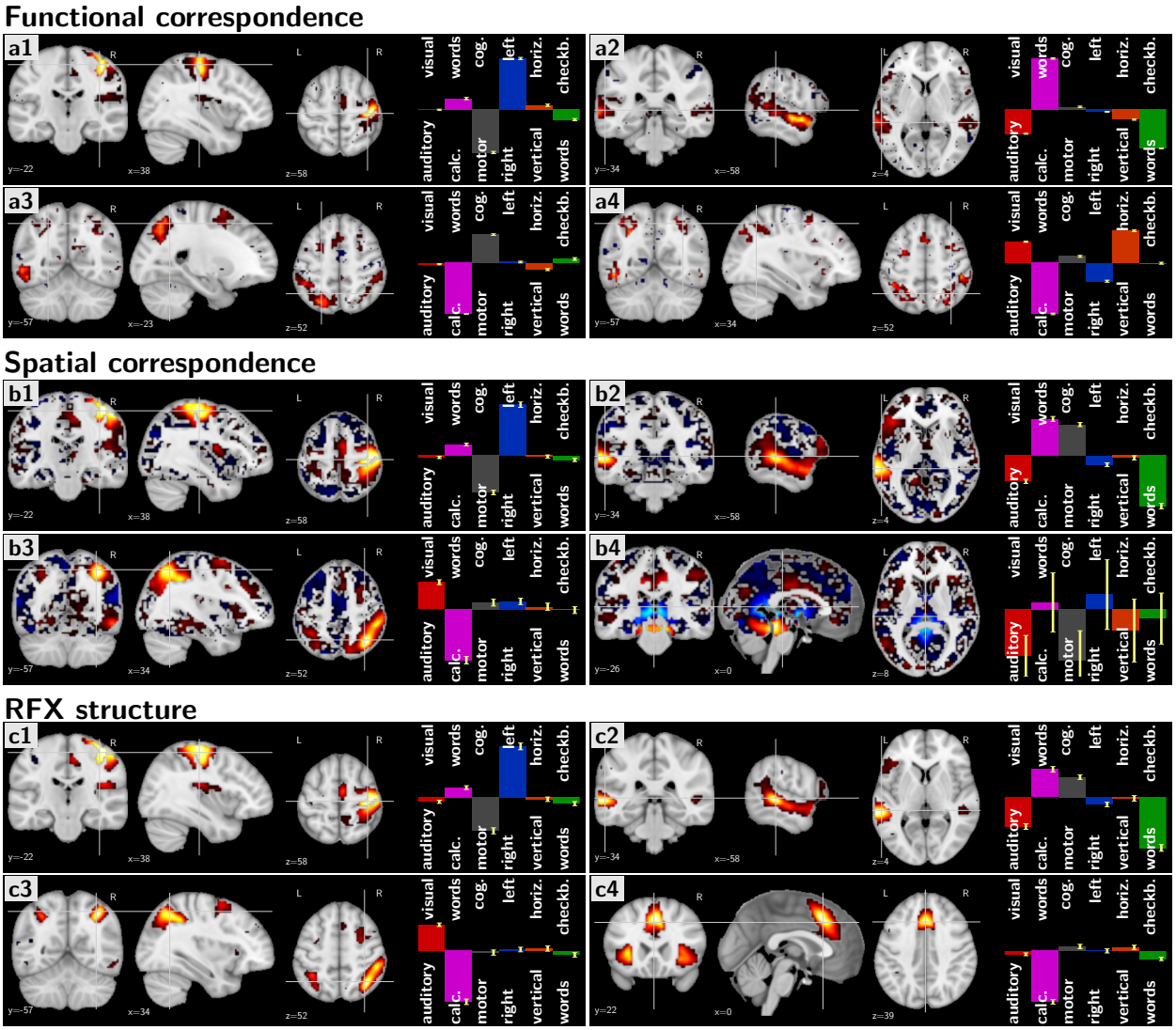


Figure 5.2: Networks learned on the localizer dataset with different strategies. Each box represents the functional network and the group-level profile as loadings on the contrasts of the study: auditory - visual, calculation - word, motor - cognition, right click - left click, vertical checkerboard - horizontal checkerboard, and words - checkerboard. The standard error across the group is displayed as a yellow bar for each loading. **a1**, **b1** and **c1** correspond to the left hand region of the motor cortex, **a2**, **b2** and **c2** to the language network, **a3**, **a4**, **b3, c3** to the dorsal-attentional network, and **c4** to a salience network. **b4** is likely a noise pattern.

network (a2, b2 and c2) and the dorsal-attentional network (a3, a4 and, b3 and c3), or a salience network (a4) [16]. We report a qualitative comparison of all the networks extracted for the different multi-subject approaches. As in the simulations, some maps learned by spatial correspondence have loadings that are not reproducible across subjects (b4 on Fig. 5.2 –note the large error bars– and on Fig. 5.3). Functional correspondence tends to mix well-known networks and produce degenerate maps. For instance, it extracts for the dorsal-attentional attentional network two components (a3 and a4) that are not well differentiated and include other regions. Indeed, the dorsal-attentional network is made of the intra-parietal sulci and the frontal eye fields and is well known for high-level visuo-spatial tasks, for instance during eye saccades. Maps a4 and a3 also outline the visual area MT (V5) and the dorsal ACC, part of respectively the visual system and the salience network. The corresponding functional profiles indeed stray away from the accepted functions of this network: a3 does not present any visual loading, while a4 shows right motor clicks and a preference for horizontal checkerboards. On the opposite, the RFX-structure approach selects only

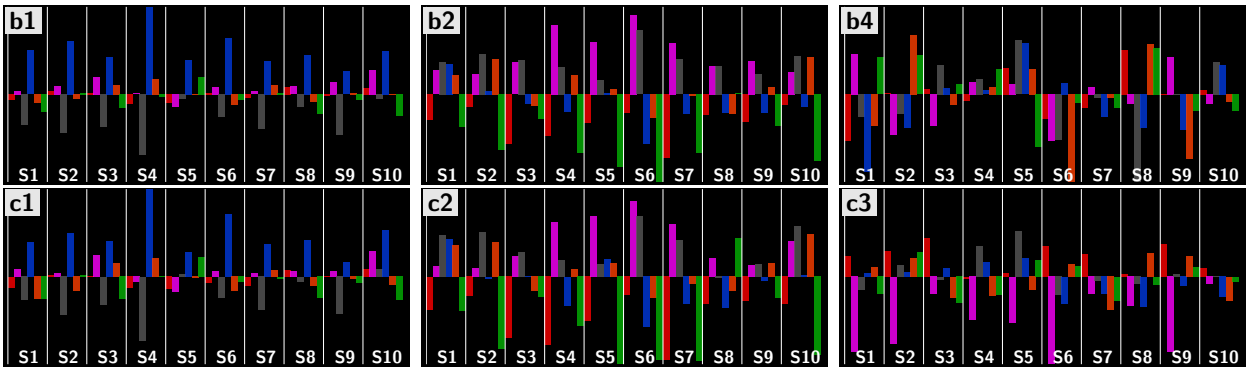


Figure 5.3: Inter-subject functional profiles $\underline{D}$ for the first 10 subjects, for spatial correspondence –top row– and RFX structure –bottom row. A white line separates subjects.

the frontal eye field and the intra-parietal sulci on the spatial map. The cognitive loadings are limited to visual and calculation tasks. While it may seem surprising to find calculation in a visuo-spatial network, this specific network has recently been reported as recruited in mental arithmetics [7]. Finally, we find that all the networks extracted by the RFX-structure approach outline known structure and have sensible cognitive loadings.

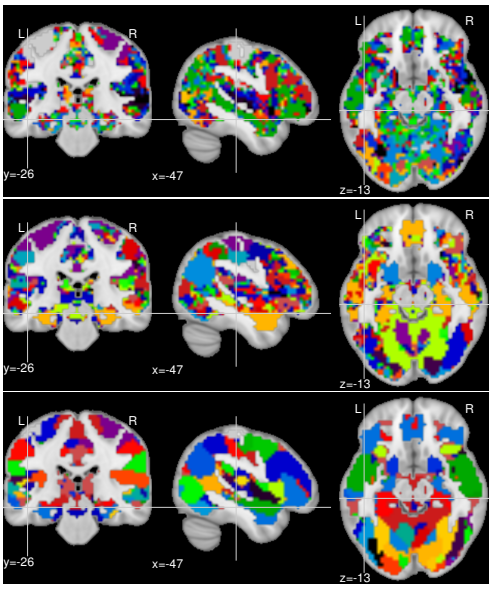


Figure 5.4: Parcellations for the different strategies. From top to bottom, functional, spatial, and RFX parcellations. The colors are random.

Towards a cognitive brain atlas To evaluate the overall spatial layout of the networks extracted we turn the decomposition in a hard assignment: we assign each voxel to the component for which it has the highest value in the spatial map. This procedure retrieves a cognitive label for each voxel and thus establishes a cognitively-informed brain parcellation. The maps extracted by functional correspondence often lack spatial structure and segment redundant regions across the different components (as with a3 and a4), as a result the corresponding parcellation appears noisy (see Fig. 5.4). The parcellations for spatial correspondence show more regularity, and even more so for the RFX-structured approach. The later gives sensible divisions of well-known parts of the cortex, such as the motor cortex, or the ventral

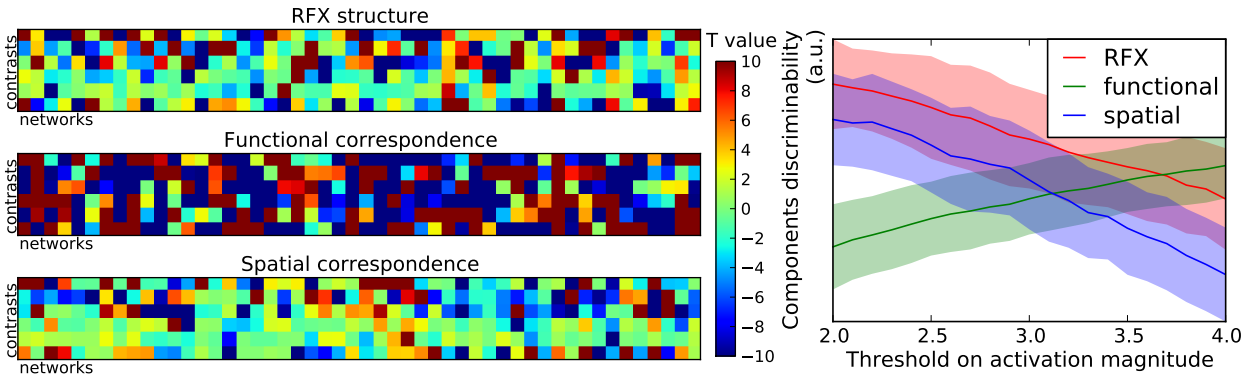


Figure 5.5: Extracted functional profiles. (Left) These profiles summarize the functional activation per network (columns) and contrast (lines) of interest through a t-value per network and contrast, across subjects. The contrasts are identical to those in Fig. 5.2. The color scale, clipped to [-10, 10], shows that the RFX model achieves an intermediate level of sensitivity. (right) The specificity of the encoding of cognitive contrasts into networks is summarized by the entropy of an assignment to negative, none or positive activation: for most thresholds the RFX model yields the most efficient encoding.

visual stream.

Functional richness of the profiles The corresponding functional profiles are summarized by computing the t-value (mean effect divided by standard error) per network and contrast, across subjects. These values, clipped to [-10, 10], are presented in Fig. 5.5(left), which shows that the RFX model achieves an intermediate level of sensitivity between spatial correspondence, that yields smaller t values, and functional correspondence that exhibits high t-values.

A way of assessing the functional significance of these decompositions is to quantify how specific the encoding of functional profiles into networks is. To do so, we label each network as showing negative, none or positive activation, by thresholding the t values, and compute the entropy of the resulting assignment. Fig 5.5 (right) presents the results for a standard range of thresholds, obtained through 100 bootstrap replications of the t values and entropy computation. In a range of values that is usable in practice (t values between 2. and 4.) the RFX model yields a more efficient encoding than the other decompositions; the spatial decomposition dominates for very low t-values while the functional decomposition outperforms the others for extremely high t values. Altogether, this suggests that the RFX model encodes efficiently the possible functional profiles, while the spatial model is more sensitive to between-subject variability and the functional model underestimates the group-level variance and thus overestimates the functional specificity of brain networks.

5.6 Conlusion

We have introduced a multi-subject model for task-induced fMRI activations that combines the principles of functional segregation and inter-subject degeneracy in a structured sparse coding problem. Technically, a major contribution of our formulation is to bound the ratio of inter-subject to intra-subject variance as it prevents extracting maps from non-reproducible variability. On a mid-sized cohort (150 subject, 6 contrasts) our model extracts a large number of brain networks that are meaningful both in terms of cognitive content and of spatial maps. Applying this approach

to larger studies should reveal richer and more specific effects. For larger cohorts, it can easily be extended to multi-level model specification, for instance in multi-centric studies, adding a center effect. An exciting direction of future research is to use this possibility to combine multiple studies in a meta analysis. Importantly, our approach is very computationally efficient: it is $O\left(n^2\right)$ in the number of subjects, and the analysis presented in this paper runs in 10 mn on a single CPU, compared to several hours for non on-line learning. It is thus applicable to mining of massive datasets. Altogether, our results provide the basis of a framework to derive a synthetic and optimized representation of large amount of multi-subject fMRI data in terms of specialized brain regions.

Bibliography

[1] P.J. Bickel, Y. Ritov, and A.B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37:1705, 2009.

[2] B. B. Biswal, M. Mennes, X. N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, et al. Toward discovery science of human brain function. *Proc Ntl Acad Sci*, 107:4734, 2010.

[3] V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar. A method for making group inferences from fMRI data using independent component analysis. *Hum Brain Mapp*, 14:140, 2001.

[4] G. Chen, E. Fedorenko, N. Kanwisher, and P. Golland. Deformation-invariant sparse coding for modeling spatial variability of functional patterns in the brain. *Machine Learning and Interpretation in Neuroimaging*, page 68, 2012.

[5] JS Damoiseaux, SARB Rombouts, F Barkhof, P Scheltens, CJ Stam, Stephen M Smith, and CF Beckmann. Consistent resting-state networks across healthy subjects. *Proceedings of the national academy of sciences*, 103(37): 13848–13853, 2006.

[6] B Efron, T Hastie, I Johnstone, and R Tibshirani. Least angle regression. *Ann. Statist.*, 32:407, 2004.

[7] A. Knops, B. Thirion, E.M. Hubbard, V. Michel, and S. Dehaene. Recruitment of an area involved in eye movements during mental arithmetic. *Science*, 324:1583, 2009.

[8] A.R. Laird, P.M. Fox, S.B. Eickhoff, et al. Behavioral interpretations of intrinsic connectivity networks. *J Cog Neurosci*, 23:4022, 2011.

[9] D. Lashkari and P. Golland. Exploratory fMRI analysis without spatial normalization. In *Information Processing in Medical Imaging*, page 398, 2009.

[10] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11:19, 2010.

[11] Maarten Mennes, Clare Kelly, Stan Colcombe, F Xavier Castellanos, and Michael P Milham. The extrinsic and intrinsic functional architectures of the human brain are not equivalent. *Cerebral Cortex*, 23(1):223–229, 2013.

[12] U. Noppeney, K.J. Friston, and C.J. Price. Degenerate neuronal systems sustaining cognitive functions. *J Anat*, 205:433, 2004.

[13] RC Petersen, PS Aisen, LA Beckett, et al. Alzheimer's disease neuroimaging initiative (ADNI) clinical characterization. *Neurology*, 74:201, 2010.

[14] P. Pinel, B. Thirion, S. Meriaux, A. Jobert, J. Serres, D. Le Bihan, J.B. Poline, and S. Dehaene. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC neuroscience*, 8:91, 2007.

[15] M.R. Sabuncu, B.D. Singer, B. Conroy, R.E. Bryan, P.J. Ramadge, and J.V. Haxby. Function-based intersubject alignment of human cortical anatomy. *Cereb Cortex*, 20:130, 2010.

[16] W.W. Seeley, V. Menon, A.F. Schatzberg, J. Keller, G.H. Glover, H. Kenna, A.L. Reiss, and M.D. Greicius. Dissociable intrinsic connectivity networks for salience processing and executive control. *J neurosci*, 27:2349, 2007.

[17] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, et al. Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci*, 106:13040, 2009.

[18] Z. Szabó, B. Póczos, and A. Lorincz. Online group-structured dictionary learning. In *CVPR*, page 2865, 2011.

[19] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.B. Poline. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Hum brain map*, 27:678, 2006.

[20] B. Thyreau, Y. Schwartz, B. Thirion, et al. Very large fMRI study using the imagen database: Sensitivity–specificity and population effect modeling in relation to the underlying anatomy. *NeuroImage*, 61:295, 2012.

[21] G. Tononi, A.R. McIntosh, D.P. Russell, and G.M. Edelman. Functional clustering: Identifying strongly interactive brain regions in neuroimaging data. *Neuroimage*, 7:133, 1998.

[22] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Inf Proc Med Imag*, page 562, 2011.

[23] Tor D Wager, Lauren Y Atlas, Martin A Lindquist, Mathieu Roy, Choong-Wan Woo, and Ethan Kross. An fMRI-based neurologic signature of physical pain. *New England Journal of Medicine*, 368(15):1388–1397, 2013.

[24] KJ Worsley, CH Liao, J. Aston, V. Petre, GH Duncan, F. Morales, and AC Evans. A general statistical analysis for fMRI data. *NeuroImage*, 15:1, 2002.

6 Learning functional atlases

Data accumulation is key to learn cognitive functions representation in the brain. The 1000 Functional Connectomes Project [2] pioneered discovery science in the neuroimaging community, by focusing on resting-state functional MRI (R-fMRI). Such an approach alleviates many of the issues related to datasharing and analysis across experiments [31], and provides an insightful view of the brain's distributed networks. R-fMRI and task-based functional MRI (T-fMRI) however outline not only the same networks [36], but also exhibit different patterns of activation [22]. Aggregating R-fMRI is therefore only addressing part of the problem, and while cognitive neuroscience traditionally uses T-fMRI to link psychological processes to brain function, a number of challenges prevent a large-scale decoding of the mental states. There indeed is a lack of frameworks that benefit from a growing accumulation of datasets. The standard analysis framework is forward inference as defined in [15, 16]: it associates a brain region with a cognitive theory implemented by a dissociation of two experimental conditions. The two conditions test for an effect of interest under an experimental paradigm, but do not guarantee that the resulting region is specific to that effect as it only reflects the cognitive theory being tested. Forward inference in a large-scale setting tends to outline broad, larger than expected regions in the brain [38], and consequently may not alone give additional insights in the context of data accumulation. A large-scale setting however opens the possibility to invert the statistical inference, and to reason from the brain activations to the mental processes, shedding light on how they interact with each other.

We have seen in Chapter V the importance of modeling brain functions as networks rather than regions, and using task fMRI rather than resting state to be able to assign labels to those networks. In this chapter, we present a supervised framework that uses the datasets presented in Chapter III, and makes it possible to learn functional atlases. The first section investigates the use of cognitive ontologies to co-analyse multiple T-fMRI datasets despite the lack of common paradigm. The second section of this chapter details the method we use to map functional labels to the brain, and the last section presents the functional atlases given by that method.

The contributions developed in this chapter have been published in:

- Y. Schwartz, G. Varoquaux, and B. Thirion, *Mapping cognitive ontologies to and from the brain*, NIPS, 2013.
- *Journal paper in preparation.*

Contents

6.1 Annotating brain maps

Price and Friston [32] argues the utility to organize cognitive processes in ontologies to enable the description of brain areas in terms of which functions they are involved in. Similarly, Poldrack [27] stresses the critical importance of the metadata describing the tasks and mental processes for scaling up the classification of mental states. This section describes some existing ontologies of cognitive processes, and how we use them to label our data.

6.1.1 Representing the cognitive space

Developing ontologies to represent mental processes is a challenging task, and there is currently no established standard in the neuroimaging community. Several initiatives however aim to solve that problem. BrainMap [20] mainly concentrates on the description of the experimental conditions that characterize an experimental paradigm. Conditions are grouped in categories that represent the stimuli, the expected responses, and the instructions given to the subjects, *e.g.,* "stimulus modality", "explicit stimulus", "explicit response". It also defines cognitive level functions called behavioral

Figure 6.1: The Cognitive Atlas lists a large number of cognitive concepts and tasks. `http://www.cognitiveatlas.org/`

domains, *e.g.,* "cognition.memory" or "cognition.language". The Cognitive Paradigm Ontology (CogPO) [39] is an extended and updated version of the BrainMap taxonomy. Figure 6.2 illustrates the decomposition of the experimental paradigms in different conditions, and their associated description. More tailored towards cognitive processes, the Cognitive Atlas [30] lists a large number of cognitive tasks and concepts, and increasingly links them together. All these approaches rely on human input to build the representations. Conversely, Neurosynth[1] [41] automatically extracts terms and associated brain locations from neuroimaging papers, to build a synthetic view of the field.

[1] http://neurosynth.org/

These approaches all have pros and cons, and ultimately are complementary. CogPO –and BrainMap– defines a formal way to specify experimental paradigms but falls short in the amount of terms it inventories, and thus in the number of studies it is able to describe. The Cognitive Atlas focuses on high level cognitive concepts that are suitable to describe contrast maps, but also sometimes appears to list redundant terms, *e.g.,* "sentence production" and "speech production". Finally Neurosynth features the larger number of terms, but their curation is an ongoing process and a lot of the terms are either redundant or not relevant. Eventually the choice of a good cognitive ontology comes down to what we want to do with it. A major challenge for this thesis is to find commonalities between studies. Despite our number of studies being large for an image database, it remains small considering that we want: i) to target a large number of mental processes to have a well-conditioned reverse inference problem, ii) several instances of each mental process to ensure the generalizability of our inference. As the description of experimental conditions is more general to all studies, regardless of the original intent of the study, we decide to mainly use terms from CogPO, and extend it where our database can benefit from more precise or high-level descriptions.

6.1.2 Representing the tasks design

Functional MRI experiments are carefully designed to balance conditions of interest with control conditions to cancel out effects related to the stimulation. As we do not want to ignore the designs, but rather leverage them in the context of a large scale inference, we introduce an additional category level for our terms, that groups together terms –or conditions– that are typically contrasted in individual studies. These new categories strongly relate to the paradigm classes from BrainMap and the tasks from the Cognitive Atlas. The categories we choose are relevant to our database, and reflect the contrasts found in the studies. They nonetheless could be modified or extended further to test other hypotheses. This hierarchy of terms can be seen as a *meta-design* that enables to co-analyse heterogeneous studies. Table 6.1 references the categories and associated terms used in this chapter.

6.2 Inferring concept-specific networks

This section introduces our method to map cognitive processes to the brain. Traditionally, researchers depend on forward inference to identify brain

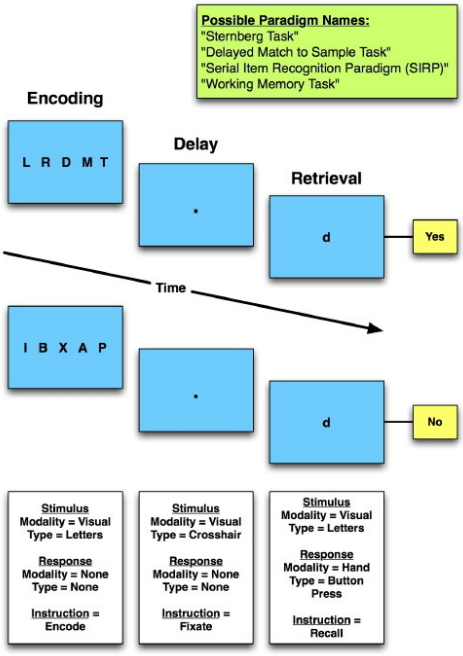


Figure 6.2: The Cognitive Paradigm Ontology describes the experimental conditions from a paradigm with a set of terms grouped in different categories. SOURCE: `http://www.cogpo.org/`

.

CogPO Categories	Task Categories	Terms
Stimulus modality	-	visual auditory
Explicit stimulus	Sounds	human voice sound
	Retinotopy	vertical checkerboard horizontal checkerboard
	Object recognition	faces places objects scramble
	Symbol recognition	visual words visual digits
Response modality	Motor - hands	left hand right hand
	Motor - feet	left foot right foot
	Arithmetics	saccades
Instructions	Arithmetics No category	calculation suppress
Cognitive Atlas terms	No category	classification language tasks
	Stop signal	inhibit failed inhibit
	Gamble	gain loss
	Flanker & Simon	congruent incongruent

Table 6.1: Terms describing the images from the database. CogPO categories are extended with new terms when needed. Cognitive Atlas terms define higher-level cognitive tasks. The task categories group terms typically used as conditions and their controls to test a hypothesis. The *stimulus modality* category stands for CogPO and task categories. Some terms do not belong to any task category and are referred as such. The *arithmetics* task category spans across the *response modality* and *instructions* CogPO categories.

patterns supporting cognitive theories. Henson [16] shows that forward inference alone is often not sufficient to conclude on the engagement of brain regions because: i) it only reflects the cognitives theories being tested, ii) it does not guarantee specificity, and iii) it cannot warrant the completeness of the inferred regions. These issues are particularly noteworthy for studies using a subtraction design, that rely on a *pure insertion* assumption [11]. Poldrack [25] points out that, to circumvent these shortcomings, researchers make use of reverse inference in an informal way: they explain the recruitment of a region by a cognitive process by relating their study with others from the literature. Henson [16] similary advocates the use of forward and reverse inference in conjunction to assess the reproducibility of an experiment. It also opens the possibility to relate cognitive processes across cognitive theories. The need of a large database is clear here, as a principled reverse inference may only be achieved with a large coverage of the cognitive space [29, 27]. Analyzing large amounts of task fMRI data usually leads to numerous challenges [2, 28]. We present a method that makes use of the terms introduced in the previous section, and enables us to use forward and reverse inference in conjunction to associate brain regions to those terms. Figure 6.3 illustrates the general idea of the method, which uses forward inference to encode the brain, *i.e.* detecting voxels responding to a term, and reverse inference to decode the brain, *i.e.* predicting the terms from the voxels.

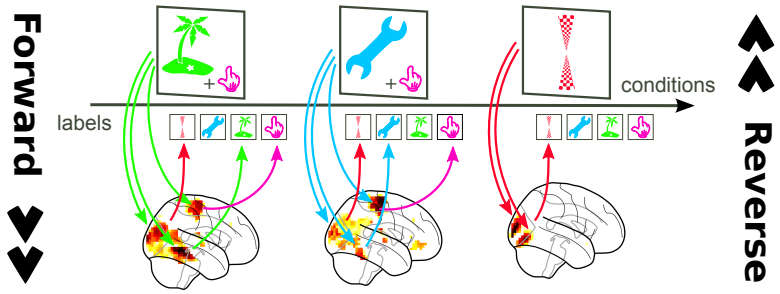


Figure 6.3: Forward inference detects voxels responding to a term, and reverse inference predicts the terms from the voxel.

.

6.2.1 Encoding the cognitive space

Terms effect Poldrack [25] formalizes forward inference as the probability of an activation given a cognitive process $\mathcal{P}(\boldsymbol{Act}|Cog)$. We assign a set of terms to each image, forming a *one-hot-encoding* of the database, *i.e.* representing the occurrence of terms by a binary design matrix. We follow the standard fMRI analysis framework and perform a General Linear Model (GLM). This gives the correlation of each separate voxel with the terms within a set of images, and enables to test for their significance. Using the GLM formulation:

$$\boldsymbol{x} = \boldsymbol{y}\boldsymbol{\beta} + \varepsilon,$$

$\boldsymbol{x}$ corresponds to the activation maps, $\boldsymbol{y}$ to the design matrix modeling the presence of terms, and $\boldsymbol{\beta}$ to the term effects. The input activation maps

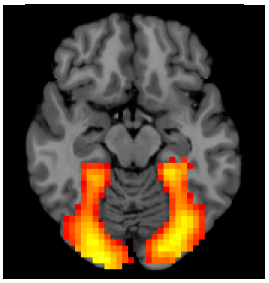


Figure 6.4: **Forward inference:** Effect map for the *places* term.

are subject-level condition versus baseline maps. The side Figure 6.4 shows the effect map for the *places* term. We will use this term along this chapter to illustrate the differences between the types of inference.

Terms contrasts The GLM estimates the responses of each voxel in respect to a combination of terms, which means the individual term effects are entailed to a degree of specificity. Individual studies contrast the $\boldsymbol{\beta}$ maps to isolate cognitive processes, e.g., a "face versus place" and a "face versus scrambled picture" contrast for a face recognition study. To disentangle the experimental factors without a too strong a priori on the control conditions, the alternative is to contrast a $\boldsymbol{\beta}$ map against all others, e.g., "face versus place and scrambled picture". We group our terms within the task categories described in Table 6.1 that contain the conditions and their controls, and proceed to compute the contrast combinations. It is important to note that this procedure is different from a 2nd level (group), or even a 3rd level analysis [33] in the sense that the term effects are estimated jointly across all studies, and the control conditions span a wider range of stimuli than typical studies.

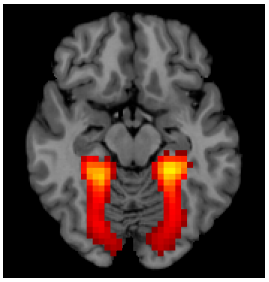


Figure 6.5: **Forward inference:** Contrast map for the *places* term, which uses faces, objects, and scrambled pictures as *controls*.

6.2.2 Decoding the cognitive space

Reverse inference Poldrack [25] formalizes reverse inference as the probability of a cognitive process given an activation $\mathcal{P}(Cog|\boldsymbol{Act})$. This is however only possible in the context of a large scale decoding framework [29], that both accounts for the distributed nature of brain networks and the variety of mental processes. In previous work, Poldrack [29, 26] tackles this question using a multi-class predictive model, the targets of the classification being separate cognitive labels. The problem we face here is slightly different, as we aim to invert the statistical inference, we predict the design matrix $\boldsymbol{y}$ denoting the entries of the design matrix in the forward model. In other terms each image may be associated with more than a single label, which in a decoding setting is called multi-label classification. This section goes through the details of our method for reverse inference, which aims to build a multi-label linear predictive model.

Feature recovery We want to build a linear model to be able to map the predictive features onto the brain. Feature recovery is the ability to recover stable and meaningful predictive features from our model. Three issues usually get in the way in fMRI multivariate analyses: the high dimensionality of the data, the local correlation of the features –voxels–, and the model selection. Varoquaux *et al.* [40] show that it is possible to come around the dimensionality and correlation problems by using sparse regression models with randomization techniques and feature clustering. This actually amounts to building an ensemble of sparse linear classifiers [7], on a set of randomized parcellations generated by a Ward agglomerative clustering algorithm combined with a resampling method. We add a cross validation procedure in the training of our ensemble in order to select the model. For each random parcellation, we keep the best model. Ensemble classifiers typically either use a voting or an averaging strategy for the fi-

nal prediction. We choose the latter to keep a linear model, in line with our brain mapping goals. We also perform a non-conservative univariate screening of the features, and keep 30% of the features. This step is primarily due to computational concerns. On the specifics of our model, we choose to use an ℓ_1-logistic regression, 5K parcels for the clustering, and a 5-fold cross validation for the model selection.

Imbalance problem The class imbalance problem is inherent to our data since mental processes are not uniformly investigated in the literature, and even more so in our database. This is a common problem for meta-analyses, known as the literature bias. There are several ways to account for class imbalance such as using resampling methods or decomposition strategies to project the classes samples into a balanced space. We choose to use a resampling method akind to bagging (Bootstrap AGGregatING), in which each classifier is given a balanced sub-sample of the whole dataset. This results in an ensemble of classifiers that retains a good coverage of the majority class but suffers less from the imbalanced class distributions.

Hierarchical decoding The previous paragraphs decribe the necessary steps to build a classifier for a single label, *i.e.* a single term, but we are in a multi-label classification setting. The usual approach to solve this kind of problem in machine learning is to train one binary classifier per label in a One versus All (OvA) scheme. The approach has successfully been used in our initial contribution [34], but in our opinion suffers from two main limitations in this context. First an OvA classification models each label separately, and by doing so misses potentially useful connections between the labels that could improve their individual prediction. Second, it ignores the experimental design of the studies from which the images are drawn: an OvA approach uses blindly all the data to learn a label, regardless of whether the images are from a study designed to expose this kind of mental process.

We introduce a new model to alleviate these shortcomings, that relies on **stacked regressions** [3]. A stacked regression model is an ensemble method that uses the linear combinations of different classifiers to improve the final prediction. The general idea of this model is to generate different predictors on the same data. The predictors can be generated through resampling methods, or merely use different underlying models (*e.g.* to combine a collection of linear and non-linear models). We *stack* the decision functions from the collection of classifiers, and use them to train a final predictor that forms a linear combination of the base models. This model has the advantage of building a linear classifier if we avoid introducing non-linearities in the ensemble classifiers. Another interesting property is that it enables to combine the predictions of OvO (One versus One) classifiers to perform multi-label classification, instead of only multi-class. Finally, this approach may be seen as a supervised dimensionality reduction method, as we condense the original space to a number of dimensions equal to the number of base classifiers in the ensemble.

Breiman [3] refers to as "black art" the choice of classifiers to form accurate combinations. As this choice is largely context dependent, we opt

for classification problems that learn a reduced representation of the brain imaging data tailored to the cognitive concepts under study. First, we stack the decisions of the OvA classifiers, that capture specific activation patterns across all tasks. This allows to relate cognitive processes across independent cognitive disciplines. Second, we build OvO classifiers by opposing terms that belong to the same task category in Table 6.1. This enables to generalize the notion of contrasts and subtraction-logic that is implicit to the majority of fMRI experiments. Finally, we build classifiers predicting the actual task categories from Table 6.1. It enables to build a hierarchical decoding framework, which combines the decisions of simpler problems (classifying the task categories), and finer grained problems (the OvO classifiers). There may be better choices of classifiers, but the final predictor weights them, and therefore mitigates the introduction of unnecessary or sub-optimal classifiers.

Finally, we learn the terms on the reduced representation with an OvA scheme, which also uses ℓ_1-penalized logistic regressions. The final output of this method is one linear classifier per term, that can be recovered by the linear combination of the coefficients of the base classifiers, with the coefficients of the final classifiers. Figure 6.6 displays the classifier's coefficients map for the "places" term: the resulting map is more specific than with forward inference, but is also noisier. Figure 6.7 summarizes this hierarchical decoding procedure.

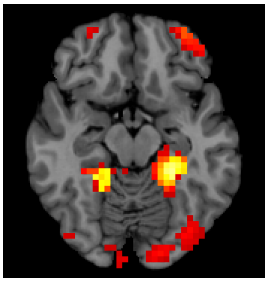


Figure 6.6: **Reverse inference:** coefficients map for the *places* term.

Cross validation We perform the classification in randomized leave-3-study out cross validation scheme. We choose to do cross-study prediction to ensure that the representation of the cognitive labels generalizes across paradigm. Failure to do so might result in over fitting the data, and learning studies idiosyncrasies. This is the first time this type of cross validation is used, as Poldrack's implementations [29, 26] relied on a leave-subject out cross validation. Considering the distribution of labels in the database, each fold enables to test only a subset of the terms. We complete 100 iterations of the cross validation to get a good estimate of the classifiers performance and variance even for the minority classes. Figure 6.8 shows the classification scores for the different labels. The precision and recall scores respectively represent the specificity and sensitivity of the classifiers. The red bars characterize the chance levels, and shows that most terms are classified above chance. We evaluate the chance levels with a classifier that generates predictions by following the class distribution in the training set. This explains why different terms have different chance levels. The green bars indicate scores with a leave-subject out cross validation, that are must higher than with a leave-study out cross validation. This confirms that the former cross validation leads to over fitting, and yet it was the approach followed in Poldrack's implementations. Figure 6.8 also provides the classification scores for a naive Bayes classifier, with the chance levels represented in red and estimated with the same method. The univariate nature of the classifier prevents it from capturing distributed patterns of activity to predict cognitive labels. As a consequence the *recall* is high –labels are often detected– but the *precision* is low because multiple cognitive labels share the same predictive features.

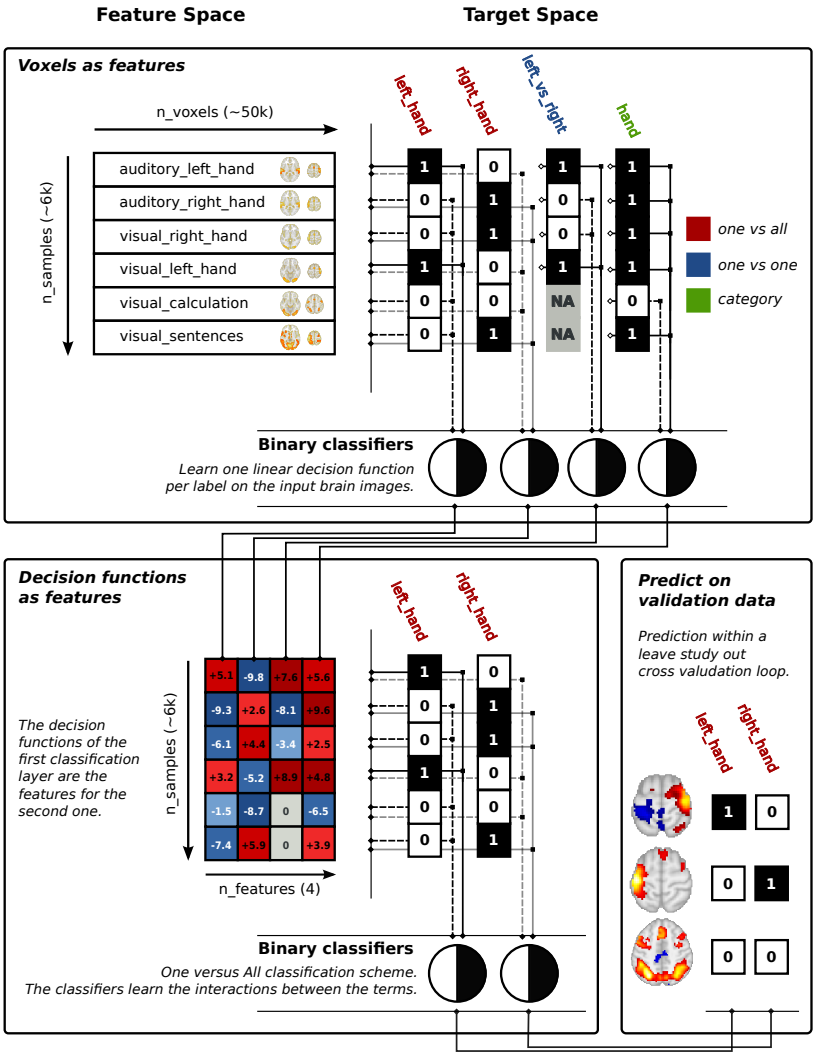
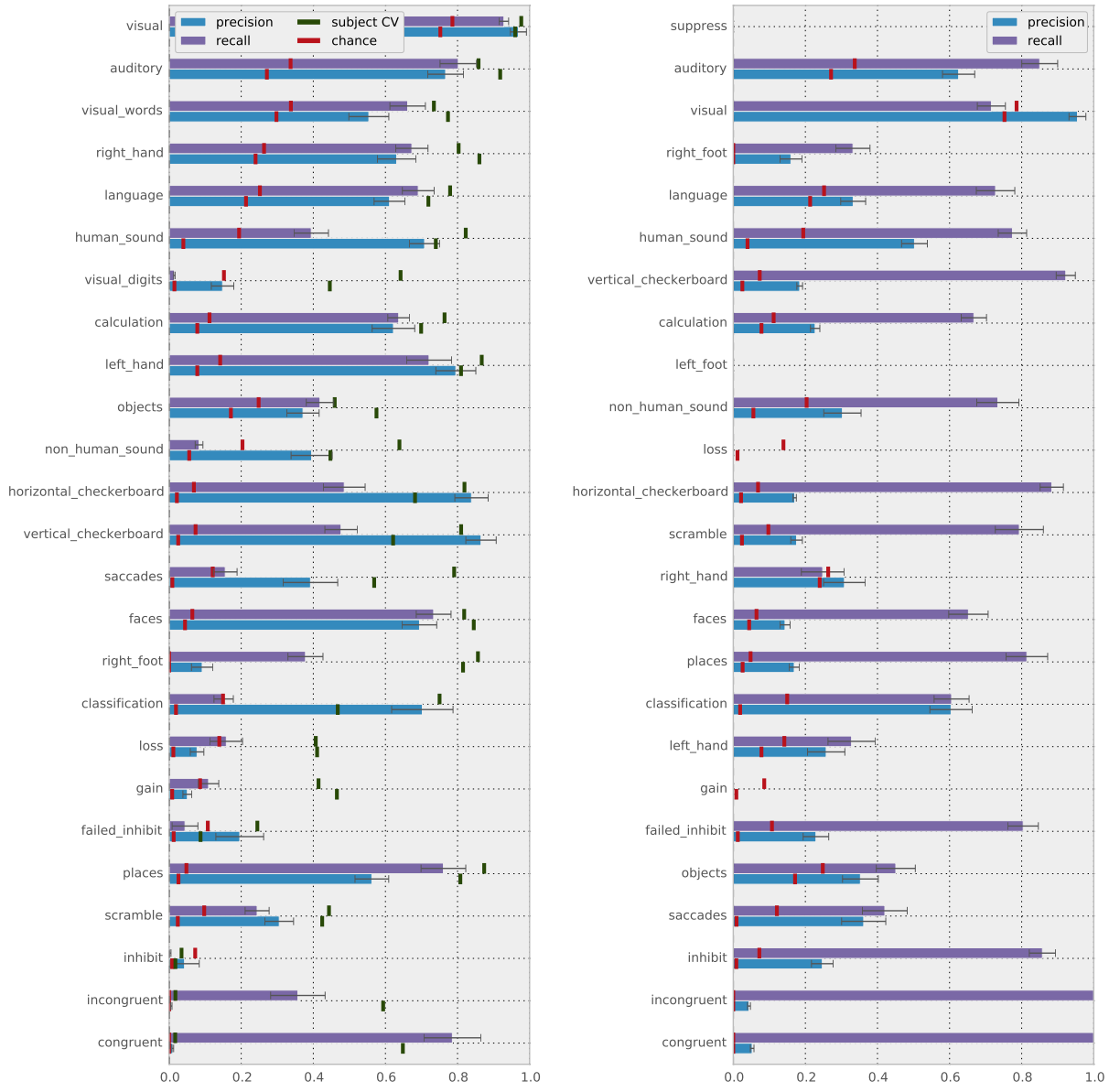


Figure 6.7: **Hierarchical decoding**. The decision functions from OvA, OvO, and task category classifiers are stacked to form a reduced feature space tailored to our problem. A second level of OvA classifiers predict the terms. Final linear classifiers may be recovered by combining the coefficients of the first and second level classifiers.

.



.

Figure 6.8: **Decoding scores**. The left figure indicate scores for the hierarchical decoding framework, the right one for a naive Bayes classifier. Precision reflects a classifier specificity and recall its sensitivity. Red bars indicate the chance levels for each term and metric, the green bars the scores obtained with a leave-subject out cross validation.

6.2.3 Mapping the cognitive space

By taking into account several cognitive concepts at the same time, reverse inference maps are more specific than the ones from forward inference, but may also capture irrelevant noise. Using both inferences in conjunction is not straightforward, as they do not perform the same statistical tests and do not have the same statistical power. As we are only interested in the common patterns between both approaches, we use a noise independent procedure to delineate those patterns. Specifically, we compute z-scores for the classifier coefficients by dividing the raw coefficients by their standard error (obtained by cross-validation). The scores' distributions are displayed in Figure 6.10, and shows the difficulty to find a scale at which to threshold forward and reverse maps to find the common patterns. For this reason, we normalize independently the forward and reverse maps. Figure 6.11 shows the z-scores' distributions after normalization. From this figure, a fair choice of threshold that yields common patterns lies between $z = 1.5$ and $z = 2$. We mask out the reverse inference maps with those from forward inference using this threshold. Figure 6.9 shows the resulting *agreement* map. Figure 6.12 summarizes all the inferences investigated in this chapter, as well as the result of using them jointly.

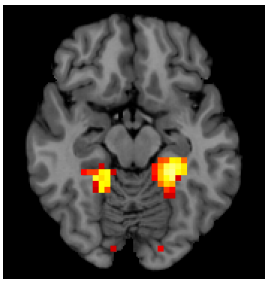


Figure 6.9: **Forward & reverse inferences:** aggreement map for the *places* term.

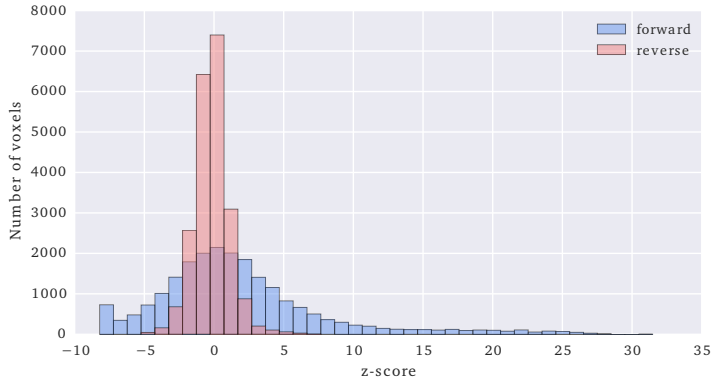


Figure 6.10: Distributions of the z-scores for forward and reverse inference for the *places* term.

.

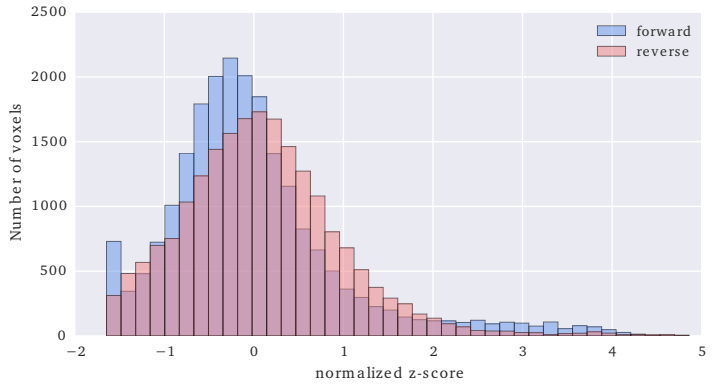
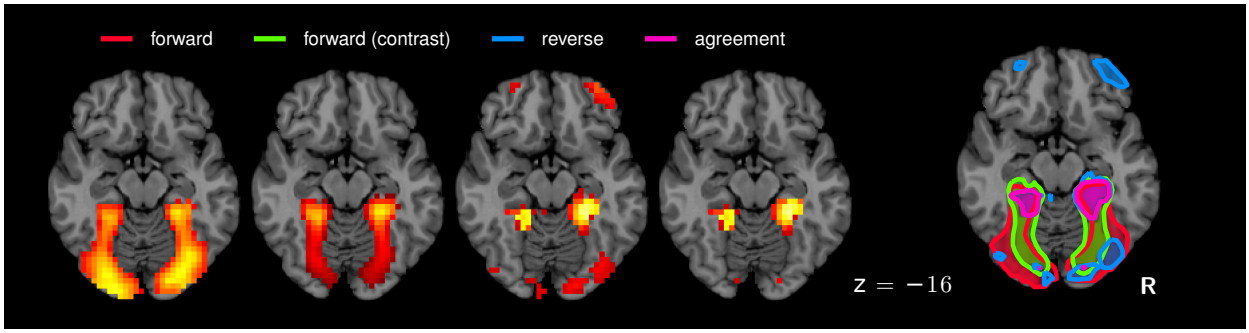


Figure 6.11: Distributions of the z-scores for forward and reverse inference for the *places* term after normalization.

.



.

Figure 6.12: **Inferences maps**. The 4 brains on the left are the forward, forward contrast, reverse, and agreement maps for the "places" term. The last brain on the right represent the overlaid inferences for the same term.

6.3 Functional atlases

This section outlines the networks exposed by the joint use of forward and reverse inference. We organize the mapping of terms in 5 *atlases* that correspond to anatomical locations: a visual atlas, a temporal atlas, a parietal atlas, a motor atlas, and a cerebellum atlas.

6.3.1 Visual atlas

Faces and places Figure 6.14 depicts the functional networks associated with visual stimuli. The "faces" and "places" labels live respectively in the FFA (Fusiform Face Area) and the PPA (Parahippocampal Place Area). While the FFA [35, 18] specializes in facial recognition and the PPA [9] in scenes encoding, reports indicate those regions are also involved in other categorical information processing [14]. The database under study however does not have such categorical stimuli to compete with the "faces" and "places" labels. Similarly, the amygdalae show involvement for "faces", which could be related to an emotional processing not otherwise modeled in the database.

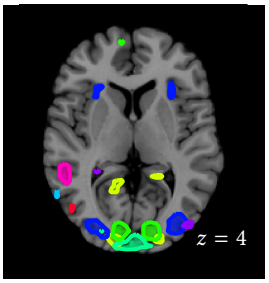


Figure 6.13: Supplementary slice of the visual atlas to show lateral occipal cortex activations for the "scramble" term (in dark blue). Full legend in Figure 6.14

.

Objects and scramble Objects opposed to scrambled pictures expose the lateral occipital complex (LOC) [21, 12], which is composed of a set of regions in the occipito-temporal cortex. Figures 6.13 and 6.14 show that the "scramble" and "object" labels delineate regions in the lateral occipital cortex and the ventral stream. Most of those regions are however labeled under the "scramble" term.

Checkerboards The "vertical and horizontal checkerboard" labels provide a basic but accurate retinal mapping. They concentrate on low-level visual processing regions such as $V1$ and $V2$.

Symbols The "digits" label is not predicted above chance and therefore does not outline any meaningful region. The "words" label exposes a region that is close to the visual word form area (VWFA), in concordance with prior art [6].

6.3.2 Temporal atlas

Auditory The "auditory" label captures the primary auditory cortex. Interestingly, the "sounds" label lies in the right planum polare (BA 38), which

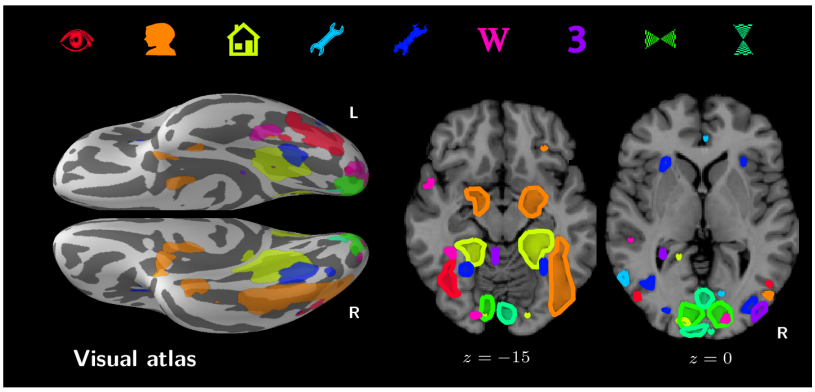


Figure 6.14: We project the labels on the surface to have a view of the ventral stream. Other views show the labels in the volume. The figure presents all the labels corresponding to visual stimui: "visual" (red eye), "faces" (orange face), "places" (yellow house), "horizontal checkerboard", "vertical checkerboard", "objects" (light blue tool), "scrambled objects" (dark blue blurred tool), "digits" (purle digit 3), and "words" (pink W letter).

.

is found to be involved in high-level music processing [5]. Dataset #40 presents a music task which could explain the recruitment of this region.

Voice area Belin *et al.* [1] identify voice-selective regions in the auditory cortex, specifically in the central part of the upper STS in both hemispheres. The human voices conditions in this study are contrasted with closely matched control conditions scrambled voices and noise. Dataset #56 features the most similar conditions in our database, by presenting Korean, French and mechanistic noise to French speakers. Other studies do not present human voices and non human sounds (such as tones, music or noise) simultaneously. The "human voice" label nonetheless depicts in Figure 6.15 a set of regions matching [1], and specific to that type of sounds.

Language The "language" label views a large distributed network in the left hemisphere composed of the anterior STS (Superior Temporal Sulcus), the posterior STS, the temporal lobe, the temporo-parietal junction, and the broca area region, which are typically reported in related work [23, 4, 10].

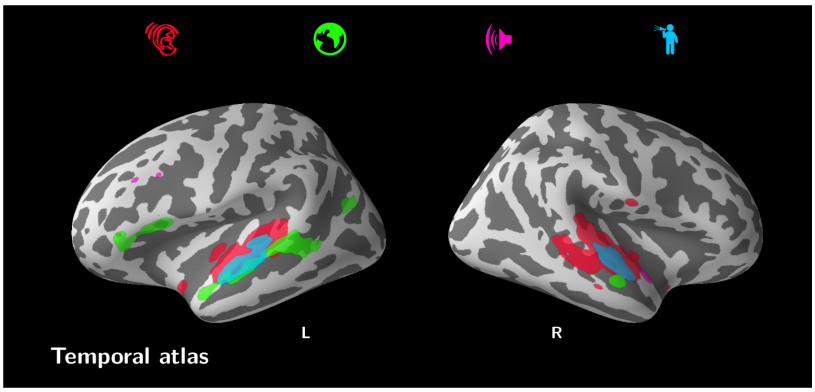


Figure 6.15: This figure presents temporal networks: "auditory" (red ear), "language" (green earth), "sound" (pink speaker), and "human voice" (light blue man).

.

6.3.3 Parietal atlas

Saccades and mental arithmetics recruit overlapping regions in the posterior parietal cortex [19]. Our method uncovers networks specific to "saccades" and "calculation" as illustrated in Figure 6.16. In particular, the "cal-

culation" network shows a strong lateralization to the left, which is consistent with previous reports [24].

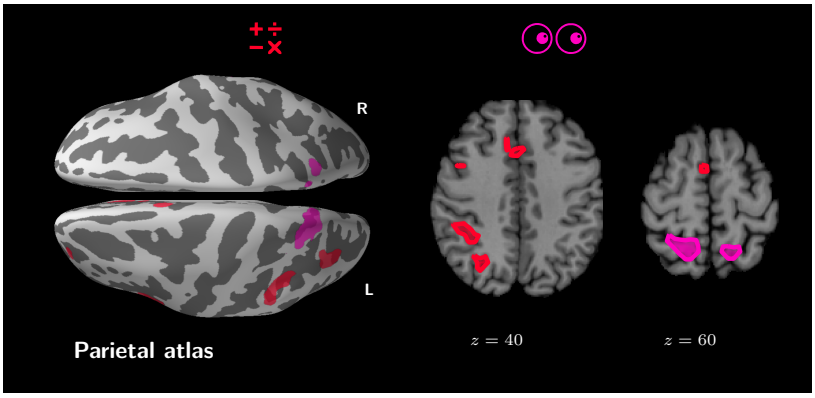


Figure 6.16: This figure presents parietal networks: "calculation" (red operators) and "saccades" (pink eyes).

.

6.3.4 Motor atlas

The motor labels in Figure 6.17 maps the hands and feets representations in the motor cortex. The cerebral hemispheres of the primary motor cortex contain the representation of the opposite side of the body part involved. Figure 6.17 also shows the motor regions in the cerebellum, which are described in more details in the cerebellum atlas section.

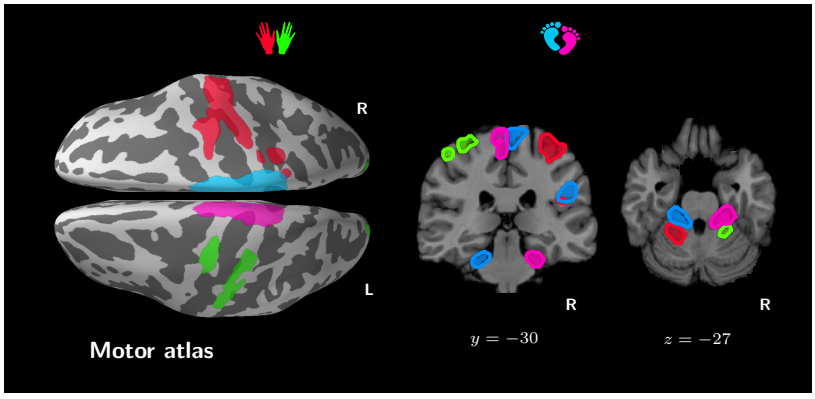


Figure 6.17: Labels related to the motor system: left hand (red), right hand (green), left foot (light blue), and right foot (pink)

.

6.3.5 Cerebellum atlas

While the cerebellar cortex engages in a range of mental processes, the extent of its functions has yet to be discovered. Previous work highlights the somatotopic organization of the cerebellum [13]. Figure 6.17 shows an analogous organization of the "hands" and "feet" functions. The crossed cerebro-cerebellar fibers pathways explains the inversion of laterality of functional areas in respect to their cortical counterparts. Other higher-level cognitive functions activate the cerebellum with the same inversion. "Language" is particularly noteworthy, and findings in [37] exhibit activations lateralized to the right lobule VI, and the Crus I and II. The language network from Figure 6.18 outlines in addition of the cortical activations a single

region in the cerebellum located in the Crus II. Finally the calculation network also involves the right cerebellar cortex in the superior medial section of the lobule VI, which is consistent with findings of working memory in the cerebellum [8]. The role of working memory in mental arithmetic is widely reported in the literature [17, 42]. These findings suggest that short-term memory is used a temporary storage of information to hold multidigits numbers, as well as to break down complex operations.

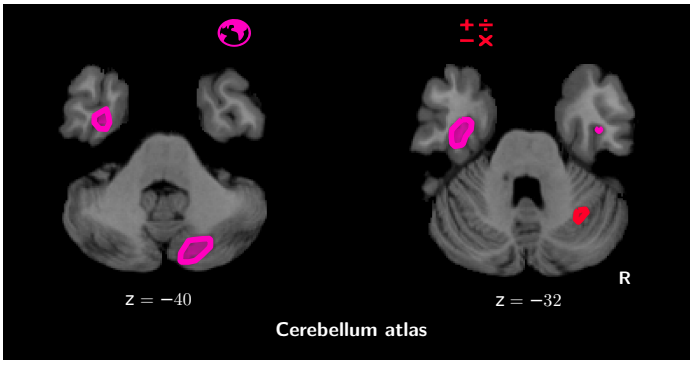


Figure 6.18: This figure shows labels, other than motor, that recruit the cerebellum: "calculation" (red operators) and "language" (pink earth).

.

6.4 Conclusion

In this chapter, we have presented a framework that is able to accumulate task fMRI data though the use of cognitives ontologies. The use of a large fMRI database opens the possibility to perform *reverse inference*, and use it in conjunction with *forward inference*. This approach, suggested by Henson, enables to go beyond functional specialization, in the sense that we do not aim to localize functions to specific brain regions, but rather identify distributed networks that may partially be shared across functions. The multivariate nature of our decoding approach is particularly beneficial to the finding of distributed networks. As a supervised approach, it can be used to assess the validity of the cognitive labels. We introduce a 2-level classification strategy, that enables to capture the specificities of the tasks paradigms. We demonstrate its ability to predict successfully low to mid-level cognitive processes, even if some higher cognitive functions still elude our model. This prediction translates to functional brain atlases that reproduce many established results. This work is the first to build a statistical model of the brain functions at such a large scale.

Bibliography

[1] Pascal Belin, Robert J Zatorre, Philippe Lafaille, Pierre Ahad, and Bruce Pike. Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–312, 2000.

[2] B. B. Biswal, M. Mennes, X. N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, et al. Toward discovery science of human brain function. *Proc Ntl Acad Sci*, 107:4734, 2010.

[3] Leo Breiman. Stacked regressions. *Machine learning*, 24(1):49–64, 1996.

[4] Jonathan Brennan, Yuval Nir, Uri Hasson, Rafael Malach, David J Heeger, and Liina Pylkkanen. Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and language*, 120(2): 163–173, 2012.

[5] Steven Brown, Michael J Martinez, Donald A Hodges, Peter T Fox, and Lawrence M Parsons. The song system of the human brain. *Cognitive Brain Research*, 20(3):363–375, 2004.

[6] Laurent Cohen and Stanislas Dehaene. Specialization within the ventral stream: the case for the visual word form area. *NeuroImage*, 22:466, 2004.

[7] Thomas G Dietterich. Ensemble methods in machine learning. In *Multiple classifier systems*, pages 1–15. Springer, 2000.

[8] Corrine Durisko and Julie A Fiez. Functional activation in the cerebellum during working memory and simple speech tasks. *Cortex*, 46(7):896–906, 2010.

[9] Russell Epstein, Alison Harris, Damian Stanley, and Nancy Kanwisher. The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, 23(1):115–125, 1999.

[10] Angela D Friederici. Broca's area and the ventral premotor cortex in language: functional differentiation and specificity. *Cortex*, 42(4):472–475, 2006.

[11] KJ Friston, CJ Price, Paul Fletcher, C Moore, RSJ Frackowiak, and RJ Dolan. The trouble with cognitive subtraction. *Neuroimage*, 4(2):97–104, 1996.

[12] Kalanit Grill-Spector. The neural basis of object perception. *Current opinion in neurobiology*, 13(2):159–166, 2003.

[13] Wolfgang Grodd, Ernst Hulsmann, Martin Lotze, Dirk Wildgruber, and Michael Erb. Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Human brain mapping*, 13(2):55–73, 2001.

[14] SJ Hanson and YO Halchenko. Brain reading using full brain support vector machines for object recognition: there is no "face" identification area. *Neural Computation*, 20:486, 2008.

[15] Richard Henson. What can functional neuroimaging tell the experimental psychologist? *The Quarterly Journal of Experimental Psychology Section A*, 58(2):193–233, 2005.

[16] Richard Henson. Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in cognitive sciences*, 10(2):64–69, 2006.

[17] Graham J Hitch. The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, 10(3): 302–323, 1978.

[18] N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, 17:4302, 1997.

[19] A. Knops, B. Thirion, E. M. Hubbard, V. Michel, and S. Dehaene. Recruitment of an area involved in eye movements during mental arithmetic. *Science*, 324:1583, 2009.

[20] Angela R Laird, Jack J Lancaster, and Peter T Fox. Brainmap. *Neuroinformatics*, 3(1):65–77, 2005.

[21] Rafael Malach, JB Reppas, RR Benson, KK Kwong, H Jiang, WA Kennedy, PJ Ledden, TJ Brady, BR Rosen, and RB Tootell. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92(18):8135–8139, 1995.

[22] Maarten Mennes, Clare Kelly, Stan Colcombe, F Xavier Castellanos, and Michael P Milham. The extrinsic and intrinsic functional architectures of the human brain are not equivalent. *Cerebral Cortex*, 23(1):223–229, 2013.

[23] C Pallier, AD Devauchelle, and S Dehaene. Cortical representation of the constituent structure of sentences. *Proc Natl Acad Sci*, 108:2522, 2011.

[24] Philippe Pinel and Stanislas Dehaene. Beyond hemispheric dominance: brain regions underlying the joint lateralization of language and arithmetic to the left hemisphere. *Journal of Cognitive Neuroscience*, 22(1):48–66, 2010.

[25] RA Poldrack. Can cognitive processes be inferred from neuroimaging data? *Trends in cognitive sciences*, 10:59, 2006.

[26] RA Poldrack, DM Barch, JP Mitchell, TD Wager, AD Wagner, JT Devlin, C Cumba, and MP Milham. Towards open sharing of task-based fMRI data: The openfMRI project (in press). *Frontiers in Neuroinformatics*.

[27] Russell A Poldrack. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron*, 72(5):692–697, 2011.

[28] Russell A Poldrack and Krzysztof J Gorgolewski. Making big data open: data sharing in neuroimaging. *Nature neuroscience*, 17(11):1510–1517, 2014.

[29] Russell A Poldrack, Yaroslav O Halchenko, and Stephen José Hanson. Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychological Science*, 20(11):1364–1372, 2009.

[30] Russell A Poldrack, Aniket Kittur, Donald Kalar, Eric Miller, Christian Seppa, Yolanda Gil, D Stott Parker, Fred W Sabb, and Robert M Bilder. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Frontiers in neuroinformatics*, 5, 2011.

[31] Jean-Baptiste Poline, Janis L Breeze, Satrajit Ghosh, Krzysztof Gorgolewski, Yaroslav O Halchenko, Michael Hanke, Christian Haselgrove, Karl G Helmer, David B Keator, Daniel S Marcus, et al. Data sharing in neuroimaging research. *Frontiers in neuroinformatics*, 6, 2012.

[32] Cathy J Price and Karl J Friston. Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, 22(3-4):262–275, 2005.

[33] G. Salimi-Khorshidi, S. M. Smith, J. R. Keltner, T. D. Wager, et al. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *Neuroimage*, 45:810, 2009.

[34] Yannick Schwartz, Bertrand Thirion, and Gaël Varoquaux. Mapping cognitive ontologies to and from the brain. *arXiv preprint arXiv:1311.3859*, 2013.

[35] Justine Sergent, Shinsuke Ohta, and Brennan Macdonald. Functional neuroanatomy of face and object processing a positron emission tomography study. *Brain*, 115(1):15–36, 1992.

[36] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, et al. Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci*, 106:13040, 2009.

[37] Catherine J Stoodley and Jeremy D Schmahmann. Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage*, 44(2):489–501, 2009.

[38] Benjamin Thyreau, Yannick Schwartz, Bertrand Thirion, Vincent Frouin, Eva Loth, Sabine Vollstädt-Klein, Tomas Paus, Eric Artiges, Patricia J Conrod, Gunter Schumann, et al. Very large fmri study using the imagen database: Sensitivity–specificity and population effect modeling in relation to the underlying anatomy. *NeuroImage*, 61(1):295–303, 2012.

[39] Jessica A Turner and Angela R Laird. The cognitive paradigm ontology: design and application. *Neuroinformatics*, 10(1):57–66, 2012.

[40] Gael Varoquaux, Alexandre Gramfort, and Bertrand Thirion. Small-sample brain mapping: sparse recovery on spatially correlated designs with randomization and clustering. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 1375–1382. ACM, 2012.

[41] T Yarkoni, RA Poldrack, TE Nichols, DC Van Essen, and TD Wager. Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8:665, 2011.

[42] Laure Zago, Mauro Pesenti, Emmanuel Mellet, Fabrice Crivello, Bernard Mazoyer, and Nathalie Tzourio-Mazoyer. Neural correlates of simple and complex mental calculation. *Neuroimage*, 13(2):314–327, 2001.

Conclusion

In this thesis, we have investigated the possibility to build upon an accumulation of fMRI data. We have reviewed the challenges to assemble a large database of images, and have proposed several novel methods to map cognitive processes to the brain. Following are the contributions of this thesis.

Database contributions Constituting a large fMRI database presents many challenges due to the complexity of the data, and the limited amount of resources available. We demonstrated the importance of a standard organization to automate the processing streams and quality assessments of the data. We also contributed to two datasharing initiatives to facilitate future projects that aim to re-use and combine existing data. Specifically, we contributed to the Neurovault project, which aims to share activation maps as a middle ground between raw data and peak coordinates. We also participated in to the BrainOmics/Localizer database, that shares both the raw data and activation maps of a single dataset featuring a large amount of stimuli.

Methodological contributions Our research focused on methods that combine multiple datasets:

- Our first contribution drew inspiration from ROI-based analysis of functional imaging data. We investigated the use of an external dataset to better condition testing on a similar task of interest. We also proposed a multivariate alternative to contrast conjunction, that learns ROIs from pairs of related fMRI experiments.

- The second contribution aimed at learning functional networks rather than ROIs. Our approach was different from other decomposition methods, in the sense that it jointly learned functional profiles that represent the networks. It therefore forgoes to manual labeling the networks with known cognitive processes.

- Our last contribution was motivated by the idea of developing a method that would map functional networks to the cognitive space. Machine learning methods generally perform better when more data is available. In our case more data also meant more diverse data, and therefore more classes (*i.e.* the cognitive processes). Handling more classes leads to a more difficult problem in terms of prediction, and mitigates the benefit of a large database. Our challenge was to frame the problem in such a way we would not suffer from the diversity of the data, and at the same time would keep a rich description of our database. Our solution was

to use a cognitive paradigm ontolgy, a describe each image with a set of cognitive labels rather than a single specific one. We used this approach both in forward and reverse inference to produce functional atlases that represent reproducible results across our datasets.

Research perspectives

Using functional profiles In our last contribution, we predicted the occurence or absence of cognitive processes. We however did not attempt to assess the difficulty of the tasks, nor represent the degree of engagement of the subjects. What we propose instead is to predict continuous variables (with a regression model) representing the involvement of a cognitive label for a task. While this would not be feasible by hand, it could be learned from the data and help the predictive models to yield more accurate decision boundaries, and therefore more accurate brain maps of cognitive labels. We propose two possible approaches to learn the functional profiles:

- By using the functional profiles from our second contribution, we can associate a certain degree of involvement to each image in the database.
- Turner et al. [1] shows the possibility to annotate fMRI experiments automatically using text mining on the papers. We could use the probability of having a given cognitive process in an experiment to define experiment-wise functional profiles.

Combining coordinate and image databases One of the most promising prospects is to combine coordinate and image databases. Image databases provide greater details on the activations, and we have shown that with a lot less studies we reach a scale comparable to coordinate databases in terms of samples. Image databases may however take some time to catch up in terms of cognitive coverage, which is a major pitfall. Given the functional profiles that represent tasks with combinations of cognitives processes and the work from Turner et al. [1] on automated annotations, we could predict unseen cognitive tasks from images. This would be a form of *zero shot learning* of tasks. Using similar ideas, we could also combine the cognitive maps learned on images with the functional profiles given by the literature to generate synthetic maps of unknown tasks.

Bibliography

[1] Matthew D Turner, Chayan Chakrabarti, Thomas B Jones, Jiawei F Xu, Peter T Fox, George F Luger, Angela R Laird, and Jessica A Turner. Automated annotation of functional imaging experiments via multi-label classification. *Frontiers in neuroscience*, 7, 2013.

A Appendix: datasets

This appendix contains the description of the datasets used in this thesis. OpenfMRI descriptions are a verbatim copy from `www.openfmri.org`, and the other datasets include a copy of the abstract of their associated paper.

A.1 Balloon Analog Risk-taking Task (ds000001)

Subjects perform the Balloon Analog Risk-taking Task in an event-related design. Note: The original highres image for sub004 was not available, so the skull-stripped version is included as highres001.nii.gz

Metadata

- Investigators: Tom Schonberg, Christopher Trepel, Craig Fox, Russell A. Poldrack
- Sample Size:16
- Scanner Type:Siemens Allegra 3T

Tasks	Experimental conditions
task001 Balloon Analogue Risk Task (BART)	cond001 Pumps fixed
	cond002 Pumps demean
	cond003 Pumps fixed real RT
	cond004 Cash fixed
	cond005 Cash demean
	cond006 Cash fixed real RT
	cond007 Explode fixed
	cond008 Explode demean
	cond009 Control pumps fixed
	cond010 Control pumps demean
	cond011 Control pumps fixed real RT

Table A.1: Tasks and experimental conditions of ds000001.

A.2 Classification learning (ds000002)

Subjects performed a classification learning task with two different problems (across different runs), using a "weather prediction" task. In one (probabilistic) problem, the labels were probabilistically related to each set of cards. In another (deterministic) problem, the labels were deterministically related to each set of cards. After learning, subjects participated in an event-related block of judgment only (no feedback) in which they were presented with stimuli from both of the training problems.

Metadata

- Investigators: Aron, A.R., Poldrack, R.A., Gluck, M.A.
- Sample Size:17
- Scanner Type:3 T Siemens Allegra MRI scanner

Tasks	Experimental conditions
task001 Probabilistic classification task	cond001 Probabilistic classification trials cond002 Feedback
task002 Deterministic classification	cond001 Deterministic classification trials cond002 Feedback
task003 Classification probe without feedback	cond001 Classification trials: Probabilistic cond002 Classification trials: Deterministic

Table A.2: Tasks and experimental conditions of ds000002.

A.3 Rhyme judgment (ds000003)

Subjects were presented with pairs of either words or pseudowords, and made rhyming judgments for each pair.

Metadata

- Investigators: Xue, G., Poldrack, R.A.
- Sample Size:13
- Scanner Type:TBA

Tasks	Experimental conditions
task001 Rhyme verification task	cond001 Word cond002 Pseudoword

Table A.3: Tasks and experimental conditions of ds000003.

A.4 Mixed-gambles task (ds000005)

Subjects were presented with mixed (gain/loss) gambles, and decided whether they would accept each gamble. No outcomes of these gambles were presented during scanning, but after the scan three gambles were selected at random and played for real money.

Metadata

- Investigators: Tom S.M., Fox C.R., Trepel C., Poldrack R.A.
- Sample Size:16
- Scanner Type:3T Siemens AG (Erlangen, Germany) Allegra MRI scanner

Tasks	Experimental conditions
task001 Mixed gambles task	cond002 Parametric gain
	cond003 Parametric loss
	cond004 Distance from indifference

Table A.4: Tasks and experimental conditions of ds000005.

A.5 Living-nonliving decision with plain or mirror-reversed text (ds000006)

Subjects performed a living-nonliving decision on items presented in either plain or mirror-reversed text. ds000006A represents the first session and ds000006B represents the second session.

Metadata

- Investigators: K Jimura, E Stover, F Cazalis, R Poldrack
- Sample Size:14
- Scanner Type:Siemens Allegra 3T

Tasks	Experimental conditions
task001 Living/nonliving judgment on mirror-reversed and plain-text words	cond001 Mirror-Switch cond002 Mirror-Repeat cond003 Plain-Switch cond004 Plain-Repeat

Table A.5: Tasks and experimental conditions of ds000006.

A.6 Stop-signal task with spoken & manual responses (ds000007)

Subjects performed a stop-signal task with one of three response types: manual response, spoken letter naming, and spoken pseudoword naming.

Metadata

- Investigators: Xue G, Aron AR, Poldrack RA
- Sample Size:20
- Scanner Type:3T Siemens Allegra MRI scanner

Tasks	Experimental conditions
task001 Stop signal task	cond001 Go trial cond002 Successful Stop Trial cond003 Unsuccessful stop trial cond004 Junk trial
task002 Stop signal task with letter naming	cond001 Go trial cond002 Successful Stop Trial cond003 Unsuccessful stop trial cond004 Junk trial
task003 Stop signal task with pseudo word naming	cond001 Go trial cond002 Successful Stop Trial cond003 Unsuccessful stop trial cond004 Junk trial

Table A.6: Tasks and experimental conditions of ds000007.

A.7 Stop-signal task with unconditional and conditional stopping (ds000008)

Subjects performed two versions of a stop signal task. In the unconditional stop-signal task, subjects are told to withhold their response whenever they hear a tone. In the conditional stop signal task, they are told to withhold their response if they hear the tone and the response is the one labeled as critical, whereas they should go ahead and respond if the response is the noncritical one. Revision history: 12/20/2012: The originally posted version of this dataset was missing some onsets for task002. The newly posted version contains the full set of onsets for all conditions. If only the onsets and model info are needed, they can be obtained by downloading the updated onsets file and untarring it in the main ds008 directory.

Metadata

- Investigators: Aron, A.R., Behrens, T.E., Frank, M., Smith, S., Poldrack, R.A.
- Sample Size:15
- Scanner Type:3T Siemens Allegra MRI scanner

Tasks	Experimental conditions
task001 Stop signal task	cond001 Go cond002 Successful stop cond003 Failed stop cond004 Junk
task002 Conditional stop signal task	cond001 Go critical cond002 Go non-critical cond003 Successful stop critical cond004 Failed stop critical cond005 Failed stop non-critical cond005 Junk

Table A.7: Tasks and experimental conditions of ds000008.

A.8 The generality of self-control (ds000009)

This study examined four different forms of self-control in a single context to determine whether multiple forms were related in a single sample of healthy adults. Participants performed four different tasks within a single scanning session.

Metadata

- Investigators: Jessica Cohen, Russell Poldrack
- Sample Size:24
- Scanner Type:Siemens Trio

Tasks	Experimental conditions
task001 Balloon Analogue Risk Task (BART)	cond001 Accept cond002 Explosion cond003 Reject
task002 Stop signal task	cond001 Go trial cond002 Successful Stop Trial cond003 Unsuccessful stop trial cond004 Junk trial
task003 Emotional regulation task	cond001 Look Negative Cue cond002 Look Neutral Cue cond003 Rating cond004 Rating - Parametric cond005 Reappraise Negative Stimulus
task004 Temporal discounting task	cond001 Easy trials cond002 Easy - parametric cond003 Hard trials cond004 Hard - parametric

Table A.8: Tasks and experimental conditions of ds000009.

A.9 Classification learning and tone-counting (ds000011)

Fourteen participants were trained on two different classification problems while they were scanned by using fMRI. Participants were trained on one problem under single-task (ST) conditions and on the other problem while performing a concurrent tone-counting task. During training, subjects learned the categories based on trial-by-trial feedback. After training, subjects received an additional block of probe trials using a mixed event-related (ER) fMRI paradigm, during which they classified items that had been trained under either ST or dual-task (DT) conditions. Tomeasure how well participants had learned under each condition, no feedback was presented during the probe block, and all items were presented under ST conditions. An additional tone-counting localizer scan presented blocks of the tone counting task (followed by a probe at the end of each block) compared to rest.

Metadata

- Investigators: Foerde, K., Knowlton, B.J., Poldrack, R.A.
- Sample Size:14
- Scanner Type:3T Siemens Allegra head-only MR scanner

Tasks	Experimental conditions
task001 Tone counting	cond001 Tone counting trials cond002 Probe
task002 Single-task weather prediction	cond001 Classification learning trials cond002 Probe
task003 Dual-task weather prediction	cond001 Classification learning trials
task004 Classification probe without feedback	cond001 Classification trials: Single task learning cond002 Classification trials: Dual task learning cond003 Junk single-task items cond004 Junk dual-task items

Table A.9: Tasks and experimental conditions of ds000011.

A.10 Classification learning and stop-signal (1 year test-retest) (ds000017)

A group of eight subjects performed two tasks (selective stop-signal and probabilistic classification) on two different occasions separated by about one year. ds000017A reflects data from timepoint 1 andds000017B reflects data from timepoint 2.

Metadata

- Investigators: Rizk-Jackson, Aron, Poldrack
- Sample Size:8
- Scanner Type:3T Siemens Allegra MRI scanner

Tasks	Experimental conditions
task001 Probabilistic classification task	cond001 Probabilistic classification trials
task002 Selective stop signal task	cond001 Go trials - critical cond002 Go trials - noncritical cond003 Successful stop cond004 Failed stop trial - critical cond005 Failed stop trial - noncritical cond006 Junk trial

Table A.10: Tasks and experimental conditions of ds000017.

A.11 Cross-language repetition priming (ds000051)

Native Spanish speakers who were proficient in English performed an abstract-concrete judgment with single Spanish or English words. Each item was repeated once, either in the same language or in the other language.

Metadata

- Investigators: Alvarez, R., Poldrack, R.A.
- Sample Size:13
- Scanner Type:Siemens Allegra

Tasks	Experimental conditions
task001 Abstract/concrete judgment: bilingual	cond001 EE-Abstract-Novel
	cond002 EE-Concrete-Novel
	cond004 ES-Concrete-Novel
	cond005 SE-Abstract-Novel
	cond006 SE-Concrete-Novel
	cond007 SS-Abstract-Novel
	cond008 SS-Concrete-Novel
	cond009 EE-Abstract-Repeat
	cond010 EE-Concrete-Repeat
	cond011 ES-Abstract-Repeat
	cond012 ES-Concrete-Repeat
	cond013 SE-Abstract-Repeat
	cond014 SE-Concrete-Repeat
	cond015 SS-Abstract-Repeat
	cond016 SS-Concrete-Repeat

Table A.11: Tasks and experimental conditions of ds000051.

A.12 Classification learning and reversal (ds000052)

Subjects performed two blocks of an event-related probabilistic classification learning task. They then performed two more blocks of the same task with the reward contingencies reversed.

Metadata

- Investigators: Poldrack, R.A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., Gluck, M.
- Sample Size:14
- Scanner Type:Siemens Allegra

Tasks	Experimental conditions
task001 Probabilistic classification task	cond001 Probabilistic classification trials: Positive feedback cond002 Probabilistic classification trials: Negative feedback
task002 Probabilistic classification task	cond001 Probabilistic reversal learning trials: Positive feedback cond002 Probabilistic reversal learning trials: Negative feedback

Table A.12: Tasks and experimental conditions of ds000052.

A.13 Simon task (ds000101)

The "NYU Simon Task" dataset comprises data collected from 21 healthy adults while they performed a rapid event-related Simon task. **Please note that all data have been uploaded regardless of quality- it is up to the user to check for data quality (movement etc). On each trial (inter-trial interval (ITI) was 2.5 seconds, with null events for jitter), a red or green box appeared on the right or left side of the screen. Participants used their left index finger to respond to the presentation of a green box, and their right index finger to respond to the presentation of a red box.In congruent trials the green box appeared on the left or the red box on the right, while in more demanding incongruent trials the green box appeared on the right and the red on the left. Subjects performed two blocks, each containing 48 congruent and 48 incongruent trials, presented in a pre-determined order (as per OptSeq), interspersed with 24 null trials (fixation only). Functional imaging data were acquired using a research dedicated Siemens Allegra 3.0 T scanner, with a standard Siemens head coil, located at the NYU Center for Brain Imaging. We obtained 151 contiguous echo planar imaging (EPI) whole-brain functional volumes (TR=2000 ms; TE=30 ms; flip angle=80, 40 slices, matrix=64x64; FOV=192 mm; acquisition voxel size=3x3x4mm) during each of the two simon task blocks. A high-resolution T1-weighted anatomical image was also acquired using a magnetization prepared gradient echo sequence (MPRAGE, TR=2500 ms; TE= 3.93 ms; TI=900 ms; flip angle=8; 176 slices, FOV=256 mm). These data have not been published previously.

Metadata

- Investigators: Kelly AMC, Milham MP
- Sample Size:21
- Scanner Type:Siemens Allegra 3.0 T scanner

Tasks	Experimental conditions
task001 Simon task	cond001 Congruent Correct
	cond002 Congruent Incorrect
	cond003 Incongruent Correct
	cond004 Incongruent incorrect

Table A.13: Tasks and experimental conditions of ds000101.

A.14 Flanker task (event-related) (ds000102)

The "NYU Slow Flanker" dataset comprises data collected from 26 healthy adults while they performed a slow event-related Eriksen Flanker task. **Please note that all data have been uploaded regardless of quality- it is up to the user to check for data quality (movement etc). On each trial (inter-trial interval (ITI) varied between 8 s and 14 s; mean ITI=12 s),participants used one of two buttons on a response pad to indicate the direction of a central arrow in an array of 5 arrows. In congruent trials the flanking arrows pointed in the same direction as the central arrow (e.g., < < < < <), while in more demanding incongruent trials the flanking arrows pointed in the opposite direction (e.g., < < > < <). Subjects performed two 5-minute blocks, each containing 12 congruent and 12 incongruent trials, presented in a pseudorandom order. Functional imaging data were acquired using a research dedicated Siemens Allegra 3.0 T scanner, with a standard Siemens head coil, located at theNYU Center for Brain Imaging. We obtained 146 contiguous echo planar imaging (EPI) whole-brain functional volumes (TR=2000 ms; TE=30 ms; flip angle=80, 40 slices, matrix=64x64; FOV=192 mm; acquisition voxel size=3x3x4mm) during each of the two flanker task blocks. A high-resolution T1-weighted anatomical image was also acquired using a magnetization prepared gradient echo sequence (MPRAGE, TR=2500 ms; TE= 3.93 ms; TI=900 ms; flip angle=8; 176 slices, FOV=256 mm). Please cite one of these papers listed below if you use these data.

Metadata

- Investigators: Kelly AMC, Uddin LQ, Biswal BB, Castellanos FX, Milham MP
- Sample Size:26
- Scanner Type:Siemens Allegra

Tasks	Experimental conditions
task001 Eriksen flanker task	cond001 Congruent Correct
	cond002 Congruent Incorrect
	cond003 Incongruent Correct
	cond004 Incongruent incorrect

Table A.14: Tasks and experimental conditions of ds000102.

A.15 Visual object recognition (ds000105)

Neural responses, as reflected in hemodynamic changes, were measured in six subjects (five female and one male) with gradient echo echoplanar imaging on a GE 3T scanner (General Electric, Milwaukee, WI) [repetition time (TR) = 2500 ms, 40 3.5-mm-thick sagittal images, field of view (FOV) = 24 cm, echo time (TE) = 30 ms, flip angle = 90] while they performed a one-back repetition detection task. High-resolution T1-weighted spoiled gradient recall (SPGR) images were obtained for each subject to provide detailed anatomy (124 1.2-mm-thick sagittal images, FOV = 24 cm). Stimuli were gray-scale images of faces, houses, cats, bottles, scissors, shoes, chairs, and nonsense patterns. The categories were chosen so that all stimuli from a given category would have the same base level name. The specific categories were selected to allow comparison with our previous studies (faces, houses, chairs, animals, and tools) or ongoing studies (shoes and bottles). Control nonsense patterns were phase-scrambled images of the intact objects. Twelve time series were obtained in each subject. Each time series began and ended with 12 s of rest and contained eight stimulus blocks of 24-s duration, one for each category, separated by 12-s intervals of rest. Stimuli were presented for 500 ms with an interstimulus interval of 1500 ms. Repetitions of meaningful stimuli were pictures of the same face or object photographed from different angles. Stimuli for each meaningful category were four images each of 12 different exemplars.

Note that the original version of the raw data that was posted prior to 10/29/2012 had one extra timepoint incorrectly added to the end of runs 1-11 for each subject. The currently posted version has been corrected.

Metadata

- Investigators: Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., Pietrini, P.
- Sample Size:6
- Scanner Type:GE 3T

Tasks	Experimental conditions
task001 Object one-back task	cond001 House
	cond002 Scrambled
	cond003 Cat
	cond004 Shoe
	cond005 Bottle
	cond006 Scissors
	cond007 Chair
	cond008 Face

Table A.15: Tasks and experimental conditions of ds000105.

A.16 Word and object processing (ds000107)

Subjects performed avisual one-back with four categories of items: written words, objects, scrambled objects and consonant letter strings.

Metadata

- Investigators: Duncan, K., Pattamadilok, C., Knierim, I., Devlin, J.
- Sample Size:49
- Scanner Type:Siemens 1.5T

Tasks	Experimental conditions
task001 Word one-back task	cond004 Consonant strings
task002 Object one-back task	cond002 Objects cond003 Scrambled

Table A.16: Tasks and experimental conditions of ds000107.

A.17 Prefrontal-Subcortical Pathways Mediating Successful Emotion Regulation (ds000108)

Although prefrontal cortex has been implicated in the cognitive regulation of emotion, the cortical-subcortical interactions that mediate this ability remain poorly understood. To address this issue, we identified a right ventrolateral prefrontal region (vlPFC) whose activity correlated with reduced negative emotional experience during cognitive reappraisal of aversive images. We then applied a pathway-mapping analysis on subcortical regions to locate mediators of the association between vlPFC activity and reappraisal success (i.e., reductions in reported emotion). Results identified two separable pathways that together explained approximately 50

Metadata

- Investigators: Wager TD, Davidson ML, Hughes BL, Lindquist MA, Ochsner KN
- Sample Size:34
- Scanner Type:1.5T GE Signa Twin Speed Excite HD scanner (GE Medical Systems)

Tasks	Experimental conditions
task001 Emotional regulation task	cond001 Look Neutral Cue
	cond002 Look Negative Cue
	cond003 Reappraise Negative Cue
	cond004 Look Neutral Stimulus
	cond005 Look Negative Stimulus
	cond006 Reappraise Negative Stimulus
	cond007 Look Neutral Rating
	cond008 Look Negative Rating
	cond009 Reappraise Negative Rating
	cond010 Look Neutral Anticipation
	cond011 Look Negative Anticipation
	cond012 Reappraise Negative Anticipation

Table A.17: Tasks and experimental conditions of ds000108.

A.18 False belief task (ds000109)

Participants read stories and answered questions that referred to either a person's false belief (mental trials) or to outdated physical representations, such as an old photograph (physical trials). Participants saw twelve stories of each type across two functional runs.

Metadata

- Investigators: Joseph M. Moran, Eshin Jolly, Jason P. Mitchell
- Sample Size:33
- Scanner Type:3T Tim Trio MRI scanner (Siemens).

Tasks	Experimental conditions
task001 False belief task	cond001 False belief story
	cond002 False belief question
	cond003 False belief photo story
	cond004 False belief photo question

Table A.18: Tasks and experimental conditions of ds000109.

A.19 Incidental encoding task (Posner Cueing Paradigm) (ds000110)

Subjects were scanned while incidentally encoding a series of visually presented real objects and greebles (meaningless objects) in a variant of the Posner cueing paradigm. Subjects covertly shifted their attention to the left or right of fixation, as cued by a centrally-presented arrow prior to item onset, and made a real object versus greeble judgment about the stimulus appearing in the cued or uncued location. Items appeared in the uncued location with a probability of .18. Subjects performed an unscanned memory test following encoding, in which they indicated their memory for old and new real objects using the following four responses: high confident old, low confident old, low confident new, high confident new. For trials in which subjects responded with one of the two old responses, a source memory judgment about the location (left or right side of the screen) of the object at study followed the recognition judgment.

Metadata

- Investigators: Melina R. Uncapher, J. Benjamin Hutchinson, Anthony D. Wagner
- Sample Size:18
- Scanner Type:3 T Signa MR scanner

Tasks	Experimental conditions
task001 Incidental encoding task	cond002 Valid HC-Hit Obj
	cond003 Valid LC-Hit Cue
	cond004 Valid LC-Hit Obj
	cond005 Valid Miss Cue
	cond006 Valid Miss Obj
	cond007 Invalid HC-Hit Cue
	cond008 Invalid HC-Hit Obj
	cond009 Invalid LC-Hit Cue
	cond010 Invalid LC-Hit Obj
	cond011 Invalid Miss Cue
	cond012 Invalid Miss Obj
	cond013 Valid Other Obj Cue
	cond015 Valid Other Greeble Cue
	cond017 Invalid Other Greeble Cue

Table A.19: Tasks and experimental conditions of ds000110.

A.20 A test-retest fMRI dataset for motor, language and spatial attention functions. (ds000114)

A test-retest dataset was acquired to validate fMRI tasks used in pre-surgical planning. In particular, five task-related fMRI time series (finger, foot and lip movement, overt verb generation, covert verb generation, overt word repetition, and landmark tasks) were used to investigate which protocols gave reliable single-subject results. Ten healthy participants in their fifties were scanned twice using an identical protocol 2-3 days apart. In addition to the fMRI sessions, high-angular resolution diffusion tensor MRI (DTI), and high-resolution 3D T1-weighted volume scans were acquired.

Metadata

- Investigators: Gorgolewski KJ, Storkey A, Bastin ME, Whittle IR, Wardlaw JM, Pernet CR
- Sample Size: 10
- Scanner Type: GE Signa HDxt 1.5T

Tasks	Experimental conditions
task001 Overt word repetition	cond001 Over word repetition
task002 Covert verb generation	cond001 Covert word generation
task003 Finger foot lips	cond001 Finger cond002 Foot cond003 Lips
task004 Overt verb generation	cond001 Overt verb generation
task005 Line bisection	cond001 Correct bisection cond002 Incorrect bisection cond003 No response task cond004 Response control cond005 No response control

Table A.20: Tasks and experimental conditions of ds000114.

A.21 Cortical processing of high-level mathematical concepts (amalric2012mathematicians)

Study featuring a visual recognition task with a focus on "digits" and "equations" stimuli, as well as a localizer task of subjects' global functions.

Metadata

- Investigators: Marie Amalric, and Stanislas Dehaene
- Sample Size: 30
- Scanner Type: Siemens 3T Trio

Tasks	Experimental conditions
task001 visual recognition task (one-back)	cond001 body
	cond002 equation
	cond003 house
	cond004 word
	cond005 number
	cond006 tool
	cond007 face
	cond008 checkerboard
task002 localizer task	cond001 horizontal checkerboard
	cond002 vertical checkerboard
	cond003 auditory right click
	cond004 auditory left click
	cond005 visual right click
	cond006 visual left click
	cond007 auditory calculation
	cond008 visual calculation
	cond009 visual sentences
	cond010 auditory sentences

Table A.21: Tasks and experimental conditions of amalric2012mathematicians.

A.22 Cortical representation of the constituent structure of sentences (devauchelle2009sentence)

Linguistic analyses suggest that sentences are not mere strings of words but possess a hierarchical structure with constituents nested inside each other. We used functional magnetic resonance imaging (fMRI) to search for the cerebral mechanisms of this theoretical construct. We hypothesized that the neural assembly that encodes a constituent grows with its size, which can be approximately indexed by the number of words it encompasses. We therefore searched for brain regions where activation increased parametrically with the size of linguistic constituents, in response to a visual stream always comprising 12 written words or pseudowords. The results isolated a network of left-hemispheric regions that could be dissociated into two major subsets. Inferior frontal and posterior temporal regions showed constituent size effects regardless of whether actual content words were present or were replaced by pseudowords (jabberwocky stimuli). This observation suggests that these areas operate autonomously of other language areas and can extract abstract syntactic frames based on function words and morphological information alone. On the other hand, regions in the temporal pole, anterior superior temporal sulcus and temporo-parietal junction showed constituent size effect only in the presence of lexico-semantic information, suggesting that they may encode semantic constituents. In several inferior frontal and superior temporal regions, activation was delayed in response to the largest constituent structures, suggesting that nested linguistic structures take increasingly longer time to be computed and that these delays can be measured with fMRI.

Metadata

- Investigators: Christophe Pallier, Anne-Dominique Devauchelle, and Stanislas Dehaene
- Sample Size: 40
- Scanner Type: Siemens 3T Trio

Tasks	Experimental conditions
task001 language task	cond001 c01
	cond002 c02
	cond003 c03
	cond004 c04
	cond005 c06
	cond006 c12
	cond007 nc3
	cond008 nc4
	cond009 motor

Table A.22: Tasks and experimental conditions of devauchelle2009sentence.

A.23 Constituent structure of sentences and music (cauvet2009muslang)

Similar experiment to A.22 but with auditory stimuli, and featuring both a language and a music task.

Metadata

- Investigators: Elodie Cauvet, Christophe Pallier
- Sample Size: 35
- Scanner Type: Siemens 3T Trio

Tasks	Experimental conditions
task001 music task	cond001 c16 music
	cond002 c08 music
	cond003 c04 music
	cond004 c02 music
	cond005 c01 music
	cond006 motor
task002 language task	cond001 c16 language
	cond002 c08 language
	cond003 c04 language
	cond004 c02 language
	cond005 c01 language
	cond006 motor

Table A.23: Tasks and experimental conditions of cauvet2009muslang.

A.24 Temporal tuning properties along the human ventral visual stream (gauthier2009resonance)

Both our environment and our behavior contain many spatiotemporal regularities. Preferential and differential tuning of neural populations to these regularities can be demonstrated by assessing rate dependence of neural responses evoked during continuous periodic stimulation. Here, we used functional magnetic resonance imaging to measure regional variations of temporal sensitivity along the human ventral visual stream. By alternating one face and one house stimulus, we combined sufficient low-level signal modulation with changes in semantic meaning and could therefore drive all tiers of visual cortex strongly enough to assess rate dependence. We found several dissociations between early visual cortex and middle- and higher-tier regions. First, there was a progressive slowing down of stimulation rates yielding peak responses along the ventral visual stream. This finding shows the width of temporal integration windows to increase at higher hierarchical levels. Next, for fixed rates, early but not higher visual cortex responses additionally depended on the length of stimulus exposure, which may indicate increased persistence of responses to short stimuli at higher hierarchical levels. Finally, attention, which was recruited by an incidental task, interacted with stimulation rate and shifted tuning peaks toward lower frequencies. Together, these findings quantify neural response properties that are likely to be operational during natural vision and that provide putative neurofunctional substrates of mechanisms that are relevant in several psychophysical phenomena as masking and the attentional blink. Moreover, they illustrate temporal constraints for translating the deployment of attention into enhanced neural responses and thereby account for lower limits of attentional dwell time.

Metadata

- Investigators: Baptiste Gauthier, Evelyn Eger, Guido Hesselmann, Anne-Lise Giraud, and Andreas Kleinschmidt
- Sample Size: 11
- Scanner Type: Siemens 3T Trio

Tasks	**Experimental conditions**
task001 continuous face house block	cond001 17ms frequency
	cond002 33ms frequency
	cond003 50ms frequency
	cond004 100ms frequency
	cond005 200ms frequency
	cond006 400ms frequency
	cond007 800ms frequency
	cond008 1600ms frequency
	cond009 3200ms frequency
	cond010 4800ms frequency
	cond011 baseline
task002 discontinuous face house block 400ms frequency	cond001 33ms duration
	cond002 50ms duration
	cond003 100ms duration
	cond004 200ms duration
	cond005 400ms duration
	cond006 baseline
task003 discontinuous face house block 800ms frequency	cond001 33ms duration
	cond002 50ms duration
	cond003 100ms duration
	cond004 400ms duration
	cond005 800ms duration
	cond006 baseline
task004 object localizer	cond001 face
	cond002 house
	cond003 object
	cond004 scramble
	cond005 baseline

Table A.24: Tasks and experimental conditions of gauthier2009resonance.

A.25 Temporal tuning properties along the human ventral visual stream (gauthier2010resonance)

See dataset A.24 for full description.

Metadata

- Investigators: Baptiste Gauthier, Evelyn Eger, Guido Hesselmann, Anne-Lise Giraud, and Andreas Kleinschmidt
- Sample Size: 13
- Scanner Type: Siemens 3T Trio

Tasks	Experimental conditions
task001 continuous face house block	cond001 50ms frequency cond002 75ms frequency cond003 100ms frequency cond004 125ms frequency cond005 150ms frequency cond006 175ms frequency cond007 200ms frequency cond008 250ms frequency cond009 400ms frequency cond010 800ms frequency cond011 baseline cond012 hits cond013 misses
task002 continuous face house block with distractor	cond001 50ms frequency cond002 75ms frequency cond003 100ms frequency cond004 125ms frequency cond005 150ms frequency cond006 175ms frequency cond007 200ms frequency cond008 250ms frequency cond009 400ms frequency cond010 800ms frequency cond011 baseline
task003 object localizer	cond001 face cond002 house cond003 object cond004 scramble cond005 baseline

Table A.25: Tasks and experimental conditions of gauthier2010resonance.

A.26 Recruitment of an Area Involved in Eye Movements During Mental Arithmetic (knops2009recruitment)

Throughout the history of mathematics, concepts of number and space have been tightly intertwined. We tested the hypothesis that cortical circuits for spatial attention contribute to mental arithmetic in humans. We trained a multivariate classifier algorithm to infer the direction of an eye movement, left or right, from the brain activation measured in the posterior parietal cortex. Without further training, the classifier then generalized to an arithmetic task. Its left versus right classification could be used to sort out subtraction versus addition trials, whether performed with symbols or with sets of dots. These findings are consistent with the suggestion that mental arithmetic co-opts parietal circuitry associated with spatial coding.

Metadata

- Investigators: André Knops, Bertrand Thirion, Edward M. Hubbard, Vincent Michel, and Dehaene Stanislas
- Sample Size: 38
- Scanner Type: Siemens 3T Trio

Tasks	Experimental conditions
task001 calculation task	cond001 operator addition
	cond002 operator subtraction
	cond003 operator color
	cond004 first operand symbolic addition
	cond005 first operand symbolic subtraction
	cond006 first operand symbolic color
	cond007 first operand non-symbolic addition
	cond008 first operand non-symbolic subtraction
	cond009 first operand non-symbolic color
	cond010 second operand symbolic addition smaller
	cond011 second operand symbolic subtraction smaller
	cond012 second operand symbolic color smaller
	cond013 second operand non-symbolic addition smaller
	cond014 second operand non-symbolic subtraction smaller
	cond015 second operand non-symbolic color smaller
	cond016 second operand symbolic addition larger
	cond017 second operand symbolic subtraction larger
	cond018 second operand symbolic color larger
	cond019 second operand non-symbolic addition larger
	cond020 second operand non-symbolic subtraction larger
	cond021 second operand non-symbolic color larger
	cond022 response
task002 saccades task	cond001 right field
	cond002 left field
task003 saccades localizer	cond001 right field
	cond002 left field

Table A.26: Tasks and experimental conditions of knops2009recruitment.

A.27 Fast reproducible identification and large-scale databasing of individual functional cognitive networks (pinel2007fast)

Although cognitive processes such as reading and calculation are associated with reproducible cerebral networks, inter-individual variability is considerable. Understanding the origins of this variability will require the elaboration of large multimodal databases compiling behavioral, anatomical, genetic and functional neuroimaging data over hundreds of subjects. With this goal in mind, we designed a simple and fast acquisition procedure based on a 5-minute functional magnetic resonance imaging (fMRI) sequence that can be run as easily and as systematically as an anatomical scan, and is therefore used in every subject undergoing fMRI in our laboratory. This protocol captures the cerebral bases of auditory and visual perception, motor actions, reading, language comprehension and mental calculation at an individual level.

Metadata

- Investigators: Philippe Pinel, Bertrand Thirion, Sébastien Meriaux, Antoinette Jobert, Julien Serres, Denis Le Bihan, Jean-Baptiste Poline, and Stanislas Dehaene
- Sample Size: 133
- Scanner Type: Siemens 3T Trio

Tasks	Experimental conditions
task001 localizer task	cond001 horizontal checkerboard
	cond002 vertical checkerboard
	cond003 auditory right click
	cond004 auditory left click
	cond005 visual right click
	cond006 visual left click
	cond007 auditory calculation
	cond008 visual calculation
	cond009 visual sentences
	cond010 auditory sentences

Table A.27: Tasks and experimental conditions of pinel2007fast.

A.28 Genetic and environmental contributions to brain activation during calculation (pinel2009twins)

Twin studies have long suggested a genetic influence on inter-individual variations in mathematical abilities, and candidate genes have been identified by genome-wide association studies. However, the localization of the brain regions under genetic influence during number manipulation is still unexplored. Here we investigated fMRI data from a group of 19 MZ (monozygotic) and 13 DZ (dizygotic) adult twin pairs, scanned during a mental calculation task. We examined both the activation and the degree of functional lateralization in regions of interest (ROIs) centered on the main activated peaks. Heritability was first investigated by comparing the respective MZ and DZ correlations. Then, genetic and environmental contributions were jointly estimated by fitting a ACE model classically used in twin studies. We found that a subset of the activated network was under genetic influence, encompassing the bilateral posterior superior parietal lobules (PSPL), the right intraparietal sulcus (IPS) and a left superior frontal region. An additional region of the left inferior parietal cortex (IPC), whose deactivation correlated with a behavioral calculation score, also presented higher similarity between MZ than between DZ twins, thus offering a plausible physiological basis for the observable inheritance of math scores. Finally, the main impact of the shared environment was found in the lateralization of activation within the intraparietal sulcus. These maps of genetic and environmental contributions provide precise candidate phenotypes for further genetic association analyses, and illuminate how genetics and education shape the development of number processing networks.

Metadata

- Investigators: Philippe Pinel, and Stanislas Dehaene
- Sample Size: 65
- Scanner Type: GE 1.5T Signa

Tasks	Experimental conditions
task001 object recognition task	cond001 words
	cond002 digit
	cond003 house
	cond004 face
	cond005 action
	cond006 tool
	cond007 scramble
task002 arithmetics and saccades task	cond001 calculation
	cond002 next number
	cond003 saccade
	cond004 junk
task003 language task	cond001 French
	cond002 Korean
	cond003 sound

Table A.28: Tasks and experimental conditions of pinel2009twins.

A.29 Principal Component Regression predicts functional responses across individuals (pinel2012archi)

Inter-subject variability is a major hurdle for neuroimaging group-level inference, as it creates complex image patterns that are not captured by standard analysis models and jeopardizes the sensitivity of statistical procedures. A solution to this problem is to model random subjects effects by using the redundant information conveyed by multiple imaging contrasts. In this paper, we introduce a novel analysis framework, where we estimate the amount of variance that is fit by a random effects subspace learned on other images; we show that a principal component regression estimator outperforms other regression models and that it fits a significant proportion (10% to 25%) of the between-subject variability. This proves for the first time that the accumulation of contrasts in each individual can provide the basis for more sensitive neuroimaging group analyzes.

Metadata

- Investigators: Bertrand Thirion, Gaël Varoquaux, Olivier Grisel, Cyril Poupon, and Philippe Pinel
- Sample Size: 76
- Scanner Type: Siemens 3T Trio

Tasks	Experimental conditions
task001 localizer task	cond001 horizontal checkerboard
	cond002 vertical checkerboard
	cond003 auditory right click
	cond004 auditory left click
	cond005 visual right click
	cond006 visual left click
	cond007 auditory calculation
	cond008 visual calculation
	cond009 visual sentences
	cond010 auditory sentences
task002 social task	cond001 intention triangle
	cond002 random triangle
	cond003 speech
	cond004 non-speech
	cond005 auditory false belief
	cond006 visual false belief
	cond007 auditory mechanistic
	cond008 visual mechanistic
task003 emotional task	cond001 trusty face
	cond002 sex face
	cond003 control face
	cond004 intention glance
	cond005 sex glance
	cond006 control glance
task004 parietal task	cond001 hand rotation
	cond002 side rotation
	cond003 saccade
	cond004 object grasp
	cond005 object orientation

Table A.29: Tasks and experimental conditions of pinel2012archi.

A.30 A Temporal Bottleneck in the Language Comprehension Network (vagharchakian2012temporal)

Humans can understand spoken or written sentences presented at extremely fast rates of ~400 wpm, far exceeding the normal speech rate (~150 wpm). How does the brain cope with speeded language? And what processing bottlenecks eventually make language incomprehensible above a certain presentation rate? We used time-resolved fMRI to probe the brain responses to spoken and written sentences presented at five compression rates, ranging from intelligible (60–100% of the natural duration) to challenging (40%) and unintelligible (20%). The results show that cortical areas differ sharply in their activation speed and amplitude. In modality-specific sensory areas, activation varies linearly with stimulus duration. However, a large modality-independent left-hemispheric language network, including the inferior frontal gyrus (pars orbitalis and triangularis) and the superior temporal sulcus, shows a remarkably time-invariant response, followed by a sudden collapse for unintelligible stimuli. Finally, linear and nonlinear responses, reflecting a greater effort as compression increases, are seen at various prefrontal and parietal sites. We show that these profiles fit with a simple model according to which the higher stages of language processing operate at a fixed speed and thus impose a temporal bottleneck on sentence comprehension. At presentation rates faster than this internal processing speed, incoming words must be buffered, and intelligibility vanishes when buffer storage and retrieval operations are saturated. Based on their temporal and amplitude profiles, buffer regions can be identified with the left inferior frontal/anterior insula, precentral cortex, and mesial frontal cortex.

Metadata

- Investigators: Laurianne Vagharchakian, Ghislaine Dehaene-Lambertz, Christophe Pallier, and Stanislas Dehaene
- Sample Size: 16
- Scanner Type: Siemens 3T Trio

Tasks	Experimental conditions
task001 visual language compression task	cond001 visual sentences 20% duration
	cond002 visual sentences 40% duration
	cond003 visual sentences 60% duration
	cond004 visual sentences 80% duration
	cond005 visual sentences 100% duration
task002 auditory language compression task	cond001 auditory sentences 20% duration
	cond002 auditory sentences 40% duration
	cond003 auditory sentences 60% duration
	cond004 auditory sentences 80% duration
	cond005 auditory sentences 100% duration

Table A.30: Tasks and experimental conditions of vagharchakian2012temporal.

A.31 A Parametric Empirical Bayesian Framework for the EEG/MEG Inverse Problem: Generative Models for Multi-Subject and Multi-Modal Integration (henson2010faces)

We review recent methodological developments within a parametric empirical Bayesian (PEB) framework for reconstructing intracranial sources of extracranial electroencephalographic (EEG) and magnetoencephalographic (MEG) data under linear Gaussian assumptions. The PEB framework offers a natural way to integrate multiple constraints (spatial priors) on this inverse problem, such as those derived from different modalities (e.g., from functional magnetic resonance imaging, fMRI) or from multiple replications (e.g., subjects). Using variations of the same basic generative model, we illustrate the application of PEB to three cases: (1) symmetric integration (fusion) of MEG and EEG; (2) asymmetric integration of MEG or EEG with fMRI, and (3) group-optimization of spatial priors across subjects. We evaluate these applications on multi-modal data acquired from 18 subjects, focusing on energy induced by face perception within a time–frequency window of 100–220ms, 8–18Hz. We show the benefits of multi-modal, multi-subject integration in terms of the model evidence and the reproducibility (over subjects) of cortical responses to faces.

Metadata

- Investigators: Richard N. Henson, Daniel G. Wakeman, Vladimir Litvak, and Karl J. Friston
- Sample Size: 16
- Scanner Type: Siemens 3T Trio

Tasks	Experimental conditions
task001 face recognition task	cond001 famous
	cond002 scrambled
	cond003 unfamiliar

Table A.31: Tasks and experimental conditions of henson2010faces.