



HAL
open science

Comparing cortical surfaces with functional magnetic resonance imaging and optimal transport : an application to decoding perceived visual semantics across individuals and species

Alexis Thual

► **To cite this version:**

Alexis Thual. Comparing cortical surfaces with functional magnetic resonance imaging and optimal transport : an application to decoding perceived visual semantics across individuals and species. Neuroscience. Sorbonne Université, 2024. English. NNT : 2024SORUS429 . tel-04906265

HAL Id: tel-04906265

<https://theses.hal.science/tel-04906265v1>

Submitted on 22 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse présentée pour l'obtention du grade de
DOCTEUR de SORBONNE UNIVERSITÉ

École doctorale
Cerveau, Cognition, Comportement (ED 158)

Spécialité
Neurosciences / Mathématiques appliquées

**Comparer les surfaces corticales
avec l'imagerie par résonance magnétique et le transport optimal**
Une application au décodage de la sémantique visuelle perçue
à travers les individus et les espèces

**Comparing cortical surfaces
with functional magnetic resonance imaging and optimal transport**
An application to decoding perceived visual semantics
across individuals and species

Par Alexis Thual

Soutenue publiquement le : **13 juin 2024**

Devant un jury composé de :

Suliann BEN HAMED, Directrice de Recherche, CNRS, ISC Marc Jeannerod

Emma ROBINSON, Senior Lecturer, King's College London

Rufin VANRULLEN, Directeur de Recherche, CNRS, CerCo

Claudia LUNGI, Chercheuse, CNRS, École Normale Supérieure

Gabriel PEYRÉ, Directeur de Recherche, CNRS, École Normale Supérieure

Wim VANDUFFEL, Directeur de Recherche, KU Leuven

Stanislas DEHAENE, Directeur de Recherche, Inserm, Collège de France, CEA

Bertrand THIRION, Directeur de Recherche, Inria Saclay

Présidente

Rapporteuse

Rapporteur

Examinatrice

Examineur

Examineur

Directeur

Co-directeur



Except where otherwise noted, this work is licensed under <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Résumé (moins de 1000 caractères)

L'imagerie non-invasive nous informe sur l'organisation fonctionnelle du cortex, mais les différences individuelles compliquent l'étude de larges populations et les comparaisons entre espèces. Nous introduisons Fused Unbalanced Gromov-Wasserstein, un nouveau modèle de Transport Optimal pour comparer les structures corticales entre individus et espèces, qui aligne l'activité fonctionnelle tout en préservant l'anatomie du cortex. Nous évaluons la pertinence des alignements calculés avec une tâche de décodage visuel. À cette fin, nous entraînons des algorithmes prédisant les perceptions visuelles sémantiques à partir de l'activité cérébrale, et montrons que l'alignement fonctionnel permet de transférer ces décodeurs à de nouveaux participants. En particulier, un décodeur entraîné chez des participants humains décode avec succès les perceptions de primates non-humains, ouvrant ainsi la voie à des comparaisons inter-espèces nouvelles que des jeux de données plus larges pourront prolonger.

Abstract (less than 1000 characters)

The rise of non-invasive neuro-imaging allows insights into brain function, yet individual differences complicate population-level conclusions and inter-species comparisons. We introduce a new method called Fused Unbalanced Gromov-Wasserstein (FUGW) that leverages Optimal Transport to compare cortical structures across individuals and species. FUGW aligns functional activity while preserving the anatomy of the cortex. We assess the relevance of computed alignments with a visual decoding task. To this end, we train algorithms predicting semantic visual perceptions from brain activity and show that functional alignment allows the transfer of these decoders to unseen participants. In particular, we show that a decoder trained in human participants can successfully decode brain activity in non-human primates. This method promises enhanced inter-species functional mapping, suggesting avenues for future research in expanding datasets and stimulus diversity.

Résumé (moins de 4000 caractères)

L'imagerie cérébrale non-invasive a permis d'acquérir des connaissances précieuses sur le fonctionnement du cerveau. En particulier, de larges groupes d'individus ont été scannés à l'aide de l'imagerie par résonance magnétique fonctionnelle (IRMf) pendant qu'ils effectuaient diverses tâches, dans l'espoir que ces données aideraient à cartographier quelles régions du cortex sont impliquées dans le traitement de stimuli ou l'exécution de tâches spécifiques. Cependant, cette entreprise est limitée par le fait que l'anatomie et l'organisation fonctionnelle du cortex diffèrent entre les individus, ce qui rend difficile l'étude de larges populations. Par ailleurs, des différences et similitudes existent également entre les primates humains et non humains, ce qui nous aide à comprendre la singularité du cerveau humain tout en complexifiant les comparaisons entre les espèces.

Nous utilisons le Transport Optimal pour modéliser ces similitudes et différences corticales entre les individus. À cette fin, nous introduisons Fused Unbalanced Gromov-Wasserstein (FUGW), un nouveau problème d'optimisation calculant des alignements entre les surfaces corticales d'individus différents en maximisant la similarité fonctionnelle entre les zones corticales appariées, tout en pénalisant les grandes déformations anatomiques du cortex. Nous montrons que, entre participants humains, les alignements calculés peuvent prédire avec succès les activations corticales liées à des stimuli n'ayant pas été utilisés pour calculer l'alignement. Cela suggère que les alignements fonctionnels obtenus sont cohérents. Nous publions un package Python open-source pour calculer ces alignements efficacement sur GPU.

Ensuite, nous étendons le package FUGW pour les données de haute résolution et fournissons une série de nouveaux solveurs. De plus, nous étudions un large spectre de configurations pour entraîner et tester ces alignements fonctionnels, en nous appuyant par exemple sur des données de visionnage de films pour calculer des alignements humains à humains, ou des données multimodales – impliquant des stimuli visuels, auditifs et tactiles – pour calculer des alignements entre humains et macaques. Cependant, ces expériences montrent l'importance d'évaluer quantitativement les alignements calculés au travers d'une nouvelle tâche pertinente pour les deux espèces.

Pour cela, nous avons recours à une tâche de décodage visuel. Ce problème complexe consiste à entraîner des modèles à prédire, à partir de leur activité corticale, quels stimuli sont actuellement vus par les participants. Nous adoptons d'abord cette approche sur des participants humains, et montrons qu'un décodeur sémantique visuel entraîné chez un individu humain peut être utilisé avec succès chez d'autres participants si leurs données sont alignées avec FUGW. Nous mesurons également que les alignements anatomiques classiques ne parviennent pas à atteindre la même performance de décodage. De plus, nous montrons que l'alignement de plusieurs participants humains per-

met de former des décodeurs plus performants grâce à un ensemble de données d'entraînement plus important. Enfin, dans un contexte inter-espèces, nous montrons qu'un décodeur cérébral entraîné chez des participants humains uniquement peut décoder avec succès la perception visuelle de primates non humains. Cela suggère que les alignements fonctionnels calculés entre les espèces contiennent des informations pertinentes concernant les similarités et les différences fonctionnelles entre les deux espèces.

Ces recherches ouvrent la voie à une cartographie fonctionnelle inter-espèces plus précise. De futurs travaux pourraient tirer parti de l'acquisition de jeux de données plus importants, nécessaires pour entraîner des décodeurs précis, et d'une formulation du problème d'alignement fonctionnel d'une manière qui ne nécessite pas que les individus alignés regardent le même ensemble de stimuli.

Abstract (less than 4000 characters)

The advent of non-invasive neuro-imaging devices has made it possible to acquire invaluable insights into how the brain works. In particular, large cohorts of patients and healthy individuals have been scanned using functional Magnetic Resonance Imaging (fMRI) while performing various tasks, hoping that this data would help map which cortical areas are involved in processing specific stimuli or performing particular tasks. However, this endeavour is limited by the fact that the anatomy and functional organisation of the cortex differ between individuals, thus making it hard to draw conclusions at the population level. Moreover, differences and similarities also exist between human and non-human primates, which, on the one hand, helps to understand the uniqueness of the human brain, and, on the other hand, makes it challenging to compare cortices across species.

We propose to tackle this issue using Optimal Transport (OT) to model cortical similarities and differences between individuals, even across species. To this end, we introduce a new OT loss, denoted as Fused Unbalanced Gromov-Wasserstein (FUGW). The underlying optimisation problem seeks to derive alignments between cortical surfaces of different individuals by maximising functional similarity between matched cortical areas, while penalising large anatomical deformations of the cortical sheet between subjects. We show that, between human participants, alignments computed with this loss can successfully predict cortical activation maps that were not used to derive the alignment, suggesting that derived functional alignments are meaningful. We release an open-source Python package implementing efficient GPU routines to derive these alignments.

In follow-up work, we extend the FUGW package to work data acquired at a high spatial resolution, and provide a series of new solvers to derive solutions to the underlying optimisation problem. Subsequently, we study a variety of different setups to train and test these functional alignments, for instance, relying on naturalistic movie-watching data to compute human-to-human alignments, or multi-modal data – involving visual, auditory, and tactile stimuli – to derive human-to-macaque alignments. However, these experiments stress the need for a way to evaluate the relevance of computed alignments using an evaluation task relevant to both species.

To this end, we design a visual decoding task that consists of training models to predict what stimuli are currently seen by participants from their brain activity. It is a complex problem, thus making it a good evaluation task to assess the relevance of computed functional alignments. Consequently, we first test this approach on human participants and show that a visual semantic decoder trained in one human subject can be used successfully in other participants if their data are functionally aligned with FUGW. Classical anatomical alignments fail to reach the same decoding performance. Moreover, we show

that aligning multiple human participants together allows one to train better-performing decoders using a larger training dataset collected from several individuals. Finally, we apply this framework in an inter-species setup and show that a brain decoder trained in human participants only can successfully decode the visual perception of non-human primates whose functional activity has been aligned to that of humans, hence suggesting that functional alignments computed across species contain relevant information regarding the functional similarities and differences between the two species.

This work paves the way towards more precise inter-species functional mapping. Future directions could involve acquiring larger datasets needed to train accurate brain decoders, and framing the functional alignment problem in a way that does not require aligned individuals to watch the same set of stimuli.

Remerciements

Je remercie Bertrand Thirion et Stanislas Dehaene de m'avoir accueilli dans leurs équipes. Je leur suis extrêmement reconnaissant de m'avoir donné l'opportunité de travailler sur des questions passionnantes et romanesques, de m'avoir soutenu dans des directions de recherche souvent risquées et d'avoir donné sans retenue leur temps pour nourrir des idées, questionner des résultats et résoudre des problèmes.

Durant ce doctorat, j'ai eu grâce à eux – et à l'environnement de recherche fertile qu'ils ont participé à créer au long cours – la liberté de participer à nombre de conférences et événements, la chance de visiter des laboratoires à la pointe de leur domaine pour m'imprégner des meilleures pratiques et faire des rencontres toujours enthousiasmantes, et enfin le privilège d'utiliser des jeux de données uniques, précieux et dont la collecte à nécessité des années d'un labeur dont d'autres que moi se sont chargés.

Mes travaux n'auraient pas été possibles sans le soutien financier du projet européen *the Human Brain Project* et les infrastructures du super ordinateur français Jean-Zay. Je mesure, qui plus est en ces temps d'incertitude politique et de prise de conscience écologique, la chance qui m'a été donnée de travailler dans de telles conditions matérielles, obtenues et entretenues par des services publics.

Je remercie Thomas Bazeille de m'avoir légué ses intuitions durement acquises et d'avoir passé de nombreux après-midi où nous avons regardé – pleins de doute et d'espoir – des cartes d'activations corticales, Maël Forcier de m'avoir entraîné à l'Institut Henri Poincaré où j'ai rencontré Huy Tran et Rémi Flamarly, mes incroyables compagnons et mentors de transport optimal, Elizabeth DuPré d'avoir partagé avec moi le meilleur café de la baie et ses instincts sur l'alignement fonctionnel avec la BOLD, Thomas Dighiero d'avoir accepté tous les cafés que je lui ai offerts, ponctué chaque phrase d'un trait d'esprit et si longuement ausculté avec moi l'alignement des cortex, Hubert Banville de m'avoir transmis l'art subtil de prédire des latents et d'avoir reçu des vers d'été comme il nous plaisait, Pierre-Louis Barbarant, mon frère de bob, de mettre ses efforts à prolonger et dépasser ces travaux.

Je remercie mes nombreux collègues et camarades de l'Inria, Neurospin et FAIR, avec une pensée particulière pour Louis Rouillard, Raphaël Meudec, Himanshu Aggarwal, Fernanda Ponce, Alexandre Blain, Cédric Allain, Thomas Chapalain, Yasmin Mzayek, Nicolas Gensollen, Hugo Richard, Pierre-Antoine Combi, Julia Linhart, Alexandre Le Bris, Houssam Zenati, Virginie Loison, Bénédicte Colnet, Matthieu Doutreligne, Charlotte Caucheteux, Philippe Ciuciu, Demian Wassermann, Thomas Moreau, Marine Le Morvan, Judith Abécassis, Nicolas Courty, Corinne Petitot, Marie Énée, Alexandre Pasquiou, Minye Zhan, Yunyun Shen, Elyes Tabbane, Pauline Martinot, Antonio Moreno, François Leroy, Andrea Adriano, Aakash Agrawal, Caroline Bévalot, Lorenzo Ciccione, Marie

Lubineau, Fosca Al Roumi, Cassandra Potier Watkins, Théo Morfoisse, Lucas Benjamin, Alexander Paunov, Audrey Maz, Tiffany Bounmy, Florent Meyniel, Christophe Pallier, Ghislaine Dehaene, Antoine Grigis, Qi Zhu, Vanna Santoro, Maryline Hévin, Yoanna Onet, Pierre Orhan, Yohann Benchetrit, Jérémy Rapin, Jean-Rémi King.

Je remercie Wim Van Duffel, Russ Poldrack, Jack Gallant, Pierre Bellec et Julie Boyle de m'avoir accueilli dans leurs équipes et d'avoir partagé avec moi leur temps et leur passion pour les neurosciences.

Je remercie Hélène et Valérie d'avoir été les premières à encourager mes balbutiements en neurosciences. J'ai une pensée pour Romain Girard, Arnaud Hirsch, Yasmine Berrada, Claire Daum, Adrien Cantérot, Raphaël Morvillier, Ahmed Ben Saoud, Corentin Dancette et Dimitri Labat qui ont écouté attentivement, plusieurs fois et avec la patience que seuls ont les vrais amis mes explications empressées sur le *decoding*, l'imagerie cérébrale et le transport optimal. Pauline, merci pour tout ce que tu as déployé pour faire advenir mes projets et les nôtres.

Durant mes recherches, il m'a souvent été donné la chance de rencontrer au petit matin des vagues clémentes et des vents favorables, des textes étourdissants et des souvenirs heureux. J'y ai trouvé beaucoup de calme et de réconfort.

Enfin, je n'ai eu de cesse pendant ces années de doctorat de penser à mes grands-parents, ma mère, mon père et Enguerrand.

Je leur dédie ce travail.

Contents

Overview	14
Notations	16
Acronyms	18
I Context	19
1 Acquisition and analysis of functional MRI data	20
1.1 Functional MRI	20
1.1.1 Principles of functional MRI	20
1.1.2 Functional MRI stimuli	25
1.1.3 Analysing functional MRI data	26
1.2 Functional brain mapping in human and non-human primates	28
1.2.1 Functional organisation of the cortex in humans	29
1.2.2 Studying the functional organisation of other species to better understand the human brain	33
1.3 Inter-individual variability and related problems	36
1.3.1 Anatomical alignment	37
1.3.2 Variability in functional images	38
1.4 Towards functional alignment	40
1.4.1 Existing voxel-to-voxel functional alignment methods	40
1.4.2 Aligning multiple individuals with functional templates	42
1.4.3 A critical perspective on current functional alignment methods	43
1.4.4 The need for open datasets, implementations and bench- marks	44
2 Introduction to Optimal Transport	46
2.1 A naive approach to functional alignment	46
2.2 Principles of Optimal Transport for Machine Learning	48
2.2.1 Historical perspective	48
2.2.2 Framing brain functional alignment as an Optimal Trans- port problem	49

2.2.3	Different classes of optimal transport problems	50
-------	---	----

II Contributions 56

3 FUGW: an optimal transport method to align cortical surfaces using anatomical and functional data 57

3.1	Introduction	57
3.2	Methods	59
3.2.1	Fused Unbalanced Gromov-Wasserstein	60
3.2.2	Approximating solutions to the FUGW problem by minimizing a lower bound	62
3.2.3	Detailed description of FUGW barycenter estimation	65
3.3	Numerical experiments	66
3.4	Results	68
3.4.1	Experiment 1 - Template anatomy	68
3.4.2	Experiment 2 - Individual anatomies	71
3.4.3	Experiment 3 - Barycenter	72
3.5	Discussion	73

4 FUGW extensions 74

4.1	Aligning human subjects with short acquisition-time fMRI training data	75
4.1.1	Introduction	75
4.1.2	Methods	76
4.1.3	Results	78
4.1.4	Discussion	78
4.2	Dependence to initialisation	79
4.3	New solvers for FUGW	81
4.4	Scaling FUGW for high spatial resolutions	81
4.4.1	A coarse-to-fine approach to scale FUGW	82
4.4.2	Adapting FUGW to use sparse matrices	83
4.4.3	Sampling high-resolution meshes uniformly	86
4.5	Precision mapping of human extrastriate visual cortex with high-resolution fMRI inter-subject alignment	87
4.5.1	Introduction	87
4.5.2	Methods	87
4.5.3	Results	88
4.5.4	Discussion	90
4.6	Comparing the convergence speed of three solvers for the FUGW problem	90
4.7	Replacing the Kullback-Leibler divergence with the squared L2-norm in the FUGW problem	92

5 Preliminary work on inter-species comparisons 95

5.1	Methodology and data collection	96
-----	---	----

5.1.1	Dataset	96
5.1.2	Alignments setup and computation	100
5.2	Results: Functional Homologies and Differences	104
5.2.1	Qualitative results	104
5.2.2	Quantitative results	104
5.3	Conclusion	108
6	Evaluating inter-subject alignment with visual decoding tasks	109
6.1	Introduction	109
6.2	Methods	112
6.2.1	Brain alignment	112
6.2.2	Decoding	114
6.2.3	Decoding and alignment setups	115
6.2.4	Datasets	115
6.2.5	Preprocessing	116
6.2.6	Hyper-parameters selection for decoders	116
6.3	Results	117
6.3.1	Within-subject prediction of visual representations from BOLD signal and retrieval of visual inputs	117
6.3.2	Out-of-subject decoding and multi-subject training	118
6.3.3	Influence of training set size and test set repetitions	120
6.4	Discussion	122
7	An application of FUGW to whole-brain inter-species functional alignment	125
7.1	Methods	125
7.1.1	MRI data	125
7.1.2	Brain decoder	127
7.1.3	Brain alignment	129
7.2	Results	131
7.2.1	Training a semantic decoder on multiple human participants	131
7.2.2	Brain decoders with functional alignment transfer from humans to macaques	131
7.2.3	Evaluating inter-species functional alignments	134
7.2.4	Choosing the best reference participant for functional alignment	138
7.3	Discussion	141
7.3.1	Limitations	141
7.3.2	Future work	141
8	Conclusions and opinions	143
8.1	Additional contributions	143
8.1.1	Contributions to the analysis of the deep-phenotyping data IBC	143

8.1.2	Software contributions to visualisation tools of neuro-imaging data	143
8.2	Limitations of the current work	145
8.2.1	Limitations of the FUGW alignment method	145
8.2.2	Limitations of the data used in the experiments	146
8.3	Opinions and future research directions	146
8.3.1	Advocating for web-based visualisation tools for neuro-imaging data	146
8.3.2	Advocating for larger deep-phenotyping fMRI datasets	147
8.3.3	Advocating for decoding-first approaches	148
III	Appendix	150
9	Supplementary material for Chapter 3 (FUGW)	151
9.1	Illustration of the alignment problem	151
9.2	Implementation details	151
9.3	Control experiments	156
9.4	Dataset description	158
10	Supplementary material for Chapter 4 (FUGW extensions)	173
10.1	Optimizing convergence speed	173
11	Supplementary material for Chapter 6 (Inter-subject visual decoding)	176
11.1	Data pre-processing	176
11.2	Retrieving images using predicted latent representations	178
11.3	Results for every combination of reference participant and left-out participant	179
11.4	Results for every type of latent representation	182
11.5	Decoding results for all setups	185
11.6	Replication on the Natural Scenes Dataset	188
12	Supplementary material for Chapter 7 (Inter-species visual decoding)	190
12.1	Additional examples of predictions	190
	Bibliography	193

Overview

The first part gives a historical and conceptual overview of the context of this thesis. Chapter 1 introduces the crucial ideas of functional Magnetic Resonance Imaging (fMRI) and how this non-invasive imaging technique can help study the cortices of human and non-human primates. It also shows how the shape and organisation of the cortex differ between individuals, even in the same species, and reviews historical *alignment* methods developed to address this issue. Chapter 2 illustrates how naive mathematical approaches to the alignment problem fail, and introduces a class of methods based on Optimal Transport (OT) to solve it.

The second part presents the main methodological contributions of this thesis. First, Chapter 3 introduces the Fused Unbalanced Gromov-Wasserstein (FUGW)¹ problem, a novel OT method to align the cortices of different individuals based on anatomical and functional data. It shows that this method successfully predicts activated cortical areas from unseen stimuli.

Secondly, Chapter 4 extends the Python package implementing FUGW to handle high-resolution data, and enriches the list of available optimisation problems and solvers. Chapter 5 builds on these extensions to present preliminary work on aligning human and non-human primate cortices. This attempt illustrates the need for an evaluation task relevant to both species to assess the quality of computed alignments.

To this end, Chapter 6 evaluates the relevance of the functional alignments between human participants on a visual decoding task. It shows that it is possible to train an algorithm to decode visual perceptions in one participant and apply it to another. Finally, Chapter 7 leverages all the previous work: it shows that one can train a decoder on human participants and successfully transfer it to non-human primates through functional alignment.

Lastly, Chapter 8 touches upon other contributions made during this thesis regarding deep-phenotyping, the development of visualisation tools for neuroimaging data through open-source projects like `brain-cockpit`² and `nilearn`³, and their deployment within the Human Brain Project⁴. Finally, it discusses the current work's limitations and future research directions.

¹<https://github.com/alexisthual/fugw>

²<https://github.com/alexisthual/brain-cockpit>

³<https://nilearn.github.io>

⁴<https://brain-cockpit.tc.humanbrainproject.eu>

Publications

Thual, A., Tran, H., Zemskova, T., Courty, N., Flamary, R., Dehaene, S., and Thirion, B. (2022). Aligning individual brains with Fused Unbalanced Gromov Wasserstein. *NeurIPS 2022*.

Thual, A., Benchetrit, Y., Geilert, F., Rapin, J., Makarov, I., Dehaene, S., Thirion, B., Banville, H., and King, J.-R. (2024a). Sample-efficient decoding of visual stimuli from fMRI through inter-individual functional alignment. Under review.

Thual, A., Wang, H., Aggarwal, H., Ponce, A. F., Depuydt, W., Zhu, Q., Vanduffel, W., Dehaene, S., and Thirion, B. (2024b). Functional alignment of MRI signal decodes visual semantics across species. In preparation.

Thirion, B., **Thual, A.**, and Pinho, A. L. (2021). From deep brain phenotyping to functional atlas. *Current Opinion in Behavioral Sciences*.

Pinho, A. L., Richard, H., Eickenberg, M., Amadon, A., Dohmatob, E., DENGHIEN, I., Torre, J. J., Shankar, S., Aggarwal, H., Ponce, A. F., **Thual, A.**, Chapalain, T., Ginisty, C., Becuwe-Desmidt, S., Roger, S., Lecomte, Y., Berland, V., Laurier, L., Joly-Testault, V., Médiouni-Cloarec, G., Doublé, C., Martins, B., Varoquaux, G., Dehaene, S., Hertz-Pannier, L., and Thirion, B. (2023). Individual Brain Charting third release, probing brain activity during Movie Watching and Retinotopic Mapping. Under review.

Thirion, B., Aggarwal, H., Ponce, A. F., Pinho, A. L., and **Thual, A.** (2023). Should one go for individual- or group-level brain parcellations? A deep-phenotyping benchmark. *Brain Structure and Function*.

Notations

\triangleq	equality by definition, for instance $\ x\ _2^2 \triangleq \langle x, x \rangle$
$\mathbf{0}_n$	$(0, \dots, 0)$ the vector of zeros in \mathbb{R}^n
$\mathbf{1}_n$	$(1, \dots, 1)$ the vector of ones in \mathbb{R}^n
I_n	the identity matrix in $\mathbb{R}^{n,n}$
$P_{\#1}$	$(\sum_j P_{i,j})_i \in \mathbb{R}^n$ the first marginal of $P \in \mathbb{R}^{n,p}$
$P_{\#2}$	$(\sum_i P_{i,j})_j \in \mathbb{R}^p$ the second marginal of $P \in \mathbb{R}^{n,p}$
$\langle \mathbf{X}, \mathbf{Y} \rangle$	$\sum_{i,j} X_{i,j} Y_{i,j}$ the generalised scalar product between $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n,p}$. We extend its definition to tensors of any matching shape
$\mathbf{X} \odot \mathbf{Y}$	$(X_{i,j} Y_{i,j})_{i,j} \in \mathbb{R}^{n,p}$ the Hadamard – i.e. element-wise – product between $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n,p}$. We extend its definition to tensors of any matching shape
$\mathbf{X} \oslash \mathbf{Y}$	$(X_{i,j} / Y_{i,j})_{i,j} \in \mathbb{R}^{n,p}$ the Hadamard – i.e. element-wise – division between $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n,p}$. We extend its definition to tensors of any matching shape
$\mathbf{X} \otimes \mathbf{Y}$	$(X_{i,j} Y_{k,l})_{i,j,k,l} \in \mathbb{R}^{n,p,q,r}$ the Kronecker – or <i>tensor</i> – product between $\mathbf{X} \in \mathbb{R}^{n,p}, \mathbf{Y} \in \mathbb{R}^{q,r}$. We extend its definition to tensors of any shape
$\mathbf{X} \oplus \mathbf{Y}$	$(X_{i,j} + Y_{k,l})_{i,j,k,l} \in \mathbb{R}^{n,p,q,r}$ the Kronecker – or <i>tensor</i> – addition between $\mathbf{X} \in \mathbb{R}^{n,p}, \mathbf{Y} \in \mathbb{R}^{q,r}$. We extend its definition to tensors of any shape
$\text{KL}(\mathbf{X} \mathbf{Y})$	$\sum_i X_i \log \frac{X_i}{Y_i}$ the Kullback-Leibler divergence between $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n,p}$, where the log and division are computed element-wise

Acronyms

ABCD Adolescent Brain Cognitive Development

BOLD Blood Oxygenation Level Dependent

DW Diffusion Weighted

FEG Frequency Encoding Gradient

FFA Fusiform Face Area

fMRI functional Magnetic Resonance Imaging

FUGW Fused Unbalanced Gromov-Wasserstein

GE-EPI Gradient Echo Echo Planar Imaging

GLM General Linear Model

HCP Human Connectome Project

HRF Hemodynamic Response Function

IBC Individual Brain Charting

ICA Independent Component Analysis

IT Inferior Temporal

MION Monocrystalline Iron Oxide Nanoparticle

MRI Magnetic Resonance Imaging

MSM Multimodal Surface Matching

MT Middle Temporal

NMR Nuclear Magnetic Resonance

OT Optimal Transport

PEG Phase Encoding Gradient

ROI Region of Interest

SRM Shared Response Model

SSG Slice Selection Gradient

STS Superior Temporal Sulcus

VFWA Visual Word Form Area

vIFC ventrolateral Frontal Cortex

Part I

Context

Chapter 1

Acquisition and analysis of functional MRI data

1.1 Functional MRI

1.1.1 Principles of functional MRI

1.1.1.1 Introduction to the physics of MRI

Magnetic Resonance Imaging (MRI) provides a non-invasive way of imaging scanned objects. To do so, it leverages Nuclear Magnetic Resonance (NMR), a physical phenomenon describing the fact that, when placed in a magnetic field, nuclei of certain atoms will absorb and re-emit radio-frequency energy. By measuring this re-emitted energy, one can gain information about the composition and behaviour of living tissues like those in the brain and reconstruct their images. One can find many introductions to the physics of MRI online. In particular, I highly recommend content from the Youtube channels *thePIRL*¹ and *Radiology Tutorials*², which published a series of videos covering a large proportion of concepts of interest in MRI physics. The following paragraphs offer a dense memo for readers already familiar with the physics of MRI.

Larmor precession Scanned objects are placed in a magnetic field \vec{B}_0 – typically on the order of a few Teslas (T) – emitted by a powerful magnet. Specific atomic nuclei, like that of Hydrogen, align with this magnetic field. When aligned with the magnetic field, these nuclei are in an unstable, high-energy state. Because this state is unstable, they can be tipped from the alignment axis by receiving a radio-frequency pulse - denoted as the *RF pulse* - on the order of tens of micro-teslas. Once tipped off the main axis, they will *precess* about it and emit a radio-frequency transverse to it. For a given nucleus, the frequency ω at

¹<https://www.youtube.com/@thepirl903>

²<https://www.youtube.com/@radiologytutorials>

which it oscillates depends on its nature: ω is proportional to its gyromagnetic ratio γ and the magnetic field strength, according to the Larmor law $\omega = \gamma \dot{B}_0$. This oscillating transverse magnetic field induces a voltage difference in coils placed around the object in the scanner. This is the signal used to reconstruct images of the object.

This signal only lasts for a limited amount of time depending on the nature of the emitting tissue: typically a few milliseconds for bones, about 100 milliseconds for white and gray matter, and up to several hundreds of milliseconds for water. Moreover, the measured signal S is the sum of all signals from each emitting particle. Thus, the total measured signal is null if the respective phases of particles are not synchronised. The theoretical duration of synchronicity between emitting particles is referred to as the *spin-spin* or *transverse relaxation time* and noted T_2 . It typically lasts a few dozen milliseconds. In practice, synchronicity decays much faster due to inhomogeneities in the magnetic field. The actual decay rate is noted T_2^* (of the order of a few milliseconds). The phenomenon in which the synchronicity between emitting particles diminishes in time is referred to as the *free induction decay (FID)*.

Finally, nuclei will eventually realign with the \vec{B}_0 field, and thus go back to their initial unstable, high-energy state. The duration of this phenomenon also depends on the nature of the tissue, and typically lasts a second. It is referred to as the *spin-lattice* or *longitudinal relaxation time* and noted T_1 . The tipping and realignment steps are repeated many times to produce images of the object scanned.

Magnetic field gradients As explained in the previous paragraph, the frequency ω at which particles oscillate is proportional to the magnetic field. If the magnetic field were uniform in space, there would be no way to determine where the emitting particles were located. To alleviate this problem, three types of gradient coils are used to modify the properties of the precessing nuclei depending on their location. First, Slice Selection Gradient (SSG) coils add spatial variations of the \vec{B}_0 field, usually along the z-axis (bottom to top). Variations are typically on the order of a few milliteslas per meter. When the gradient is active, particles located at different heights will react to different frequencies ω of the RF pulse. Therefore, if a single frequency ω is emitted, only particles of a given slice will emit a coherent signal that can be measured. SSG coils are usually switched on simultaneously with the RF pulse.

Then, the same principle is employed in the selected slice to modify the frequency of precession of particles along the x-axis (left to right), using Frequency Encoding Gradient (FEG) coils. Finally, Phase Encoding Gradient (PEG) coils will produce a gradient along the y-axis (front to back) so that some nuclei will precess faster and some slower. Frequency-encoding and phase-encoding coils are typically switched on during a few milliseconds, and only after the RF pulse was sent.

It is important to note that changes in magnetic gradients are not instanta-

neous: the speed of change is measured by the *slew rate*, which is usually on the order of magnitude of 10-200 mT / m / ms.

Image reconstruction with inverse Fourier transform The slice selection coil ensures that only particles in a given slice are excited by the RF pulse. The frequency-encoding and phase-encoding coils will make particles in the slice oscillate at different frequencies and phases depending on their (x, y) location. One can show that the magnetisation $M(x, y)$ emitted by particles in the slice is now multiplied by a complex sinusoidal function $e^{i\phi(x,y)}$. The key idea behind modern MRI is to measure the overall signal S for different, chosen values of $\phi(x, y)$, as they form the discrete Fourier transform \hat{M} of the signal of interest $M(x, y)$. Therefore, one does not observe $M(x, y)$ directly, but rather its Fourier transform $\hat{M}(k_x, k_y)$, where k_x and k_y are coordinates in the so-called *k-space*. Note that with our notations, $\phi(x, y) \triangleq k_x x + k_y y$ and $S(k_x, k_y) = \int \int M(x, y) e^{i\phi(x,y)} dx dy$. Importantly, the value of k_x and k_y is controlled by the encoding coils – it is proportional to the integral of the frequency- and phase-encoding coils in time. One can thus navigate the k-space and measure the complex value $\hat{M}(k_x, k_y)$ at different coordinates (k_x, k_y) . Finally, M can be reconstructed by applying the inverse Fourier transform to \hat{M} , which yields an image of the slice scanned. Repeating this process for many different slices will allow one to reconstruct the 3D image of the object scanned. As a note, it can be helpful to think of images being built up from a combination of spatial frequencies, with low frequencies corresponding to large structures and high frequencies corresponding to fine details. The center of k-space – i.e. values of k_x and k_y close to zero – corresponds to low frequencies, and the edges of k-space correspond to high frequencies. We encourage the interested reader to look at the FSL course on MRI physics³ for more information.

MRI sequences Navigating the k-space – i.e. choosing the dynamics of gradient coils – can be done in many different ways. Different setups are referred to as *MRI sequences*. A naive MRI imaging approach would send a single RF pulse, turn on the frequency- and phase-encoding coils for a brief moment, measure the signal emitted by the object scanned for the underlying (k_x, k_y) coordinate, and wait for particles to realign with the \vec{B}_0 field before repeating the process for different values of (k_x, k_y) . However, as T_1 durations are typically on the order of a second, this would lead to very slow imaging. Instead, researchers have developed MRI sequences making complex use of gradient coils so as to acquire many different phases ϕ without having to wait for longitudinal relaxation. Moreover, the RF pulse does not need to align particles orthogonally to the \vec{B}_0 field: it can tip them by any given *flip angle* α . Smaller values of α will result in shorter longitudinal relaxation times. The best value of α is that which gives maximal transverse signal while yielding short acquisition times. Consequently, different sequences have been developed. For dynamical images of the brain,

³https://open.win.ox.ac.uk/pages/fslcourse/lectures/additional/2019/fsl_introMRI.pdf

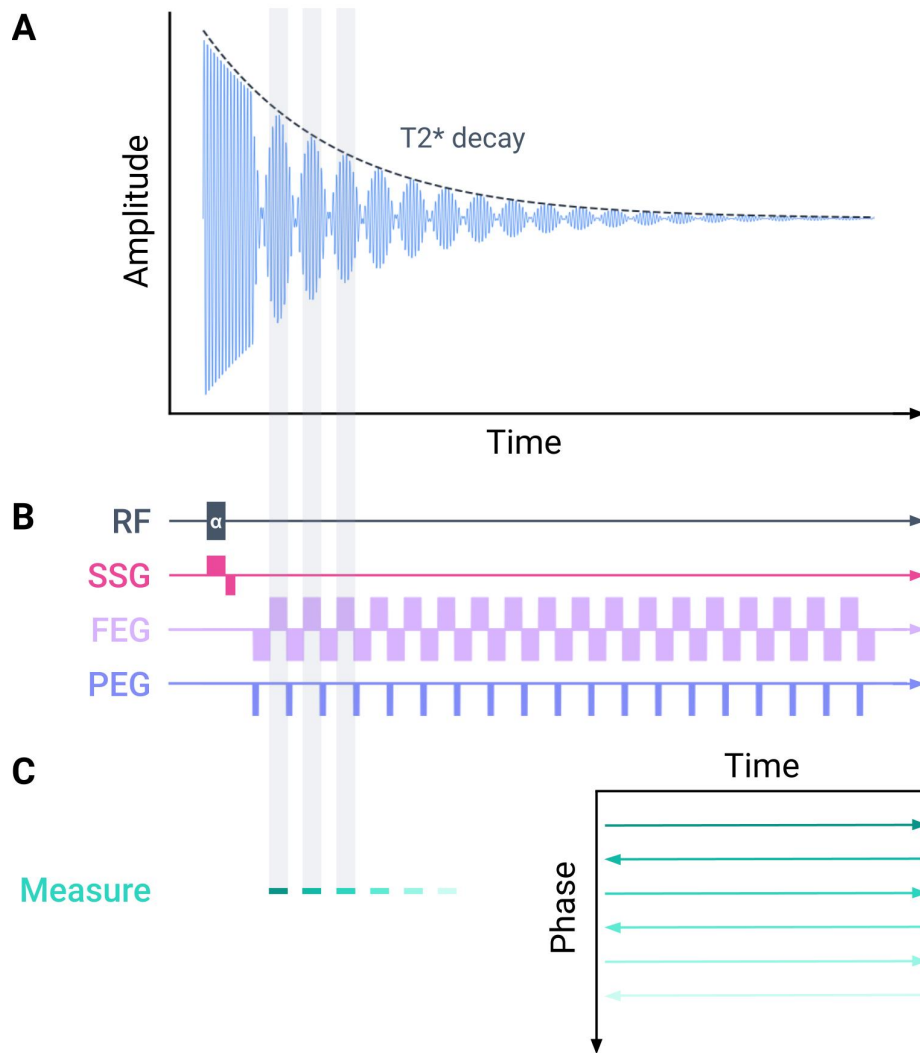


Figure 1.1: Schematic view of the GE-EPI MRI sequence

A. Coils of the MRI measure the amplitude of the signal emitted by the object scanned. This signal is the sum of all signals emitted by particles in the object. Its amplitude decays at the rate of T_2^* . Its shape is affected by gradient coils. **B.** The RF pulse tips particles off the \vec{B}_0 field by an angle α . The slice selection gradient coil ensures that only particles in a given slice are affected. The frequency-encoding and phase-encoding gradients encode the (x, y) coordinates of particles in the slice. **C.** The overall signal is measured many times when the signal's amplitude increases due to gradient echo. An inverse Fourier transform run on all samples will indicate what frequencies and phases have yielded the observed signal. The k-space (right) is travelled in a so-called cartesian trajectory indicated by green arrows.

Adapted from the MRI Physics Course #21 from Radiology Tutorials.

like those studied in this thesis, the most commonly used sequence is that of Gradient Echo Echo Planar Imaging (GE-EPI) displayed in Figure 1.1. Finally, acquiring signal from multiple slices simultaneously using multi-band sequences is possible, significantly reducing the time needed to acquire images.

1.1.1.2 Acquiring MRI images of the brain

T1- and T2-weighted images Researchers can rely on either the T_1 and T_2^* relaxation times to gain information about the composition of tissues in the scanned object. At a given time t , the signal a tissue emits depends on the relaxation times T_1 and T_2^* of this tissue. T_1 , typically longer than T_2^* , is used to contrast between tissues and reconstruct anatomical images of the brain. These are called T_1 -weighted images. On the other hand, T_2^* is much shorter and is typically used to study the dynamics of the brain. These are called T_2^* -weighted images, or sometimes abusively T_2 -weighted images.

BOLD In particular, MRI can measure changes in blood flow in the brain, which correlate with neural activity. More precisely, neurons consume energy from glucose and oxygen. When active, they consume more energy and need an additional supply of oxygen provided by the blood. Depending on the activity of neurons, surrounding blood vessels will exhibit different levels of oxy- and deoxyhemoglobin. [Ogawa et al., 1990](#) showed that oxy- and deoxyhemoglobin have different magnetic properties, allowing one to distinguish them using MRI. It is the principle of Blood Oxygenation Level Dependent (BOLD) imaging, and blood oxygenation is a proxy for neural activity.

It is important to note that modifications of oxy- and deoxyhemoglobin concentration are not instantaneous. After neurons have been activated, it takes a few seconds for the oxygen level in the blood to peak before it decreases back to normal. The concentration's evolution due to neuronal activity is called the Hemodynamic Response Function (HRF). Interestingly, the shape of the HRF varies across areas of the cortex of a given individual, with aging, and across participants in general.

Contrast agents BOLD imaging is routinely used to image the dynamics of the human brain. However, it may not be sensitive enough to probe the dynamics of smaller systems, like those of monkeys. Consequently, researchers have resorted to using contrast agents. These agents are injected into the bloodstream and modify the blood's magnetic properties for a limited time, allowing one to measure the dynamics of smaller systems. In particular, the authors of [Leite et al., 2002](#) show that injecting macaques with Monocrystalline Iron Oxide Nanoparticle (MION) can lead to a 3-fold increase in signal changes of functional images compared to BOLD. Note that the response function associated with contrast agents can differ greatly from that of BOLD, as illustrated in Figure 1.2.

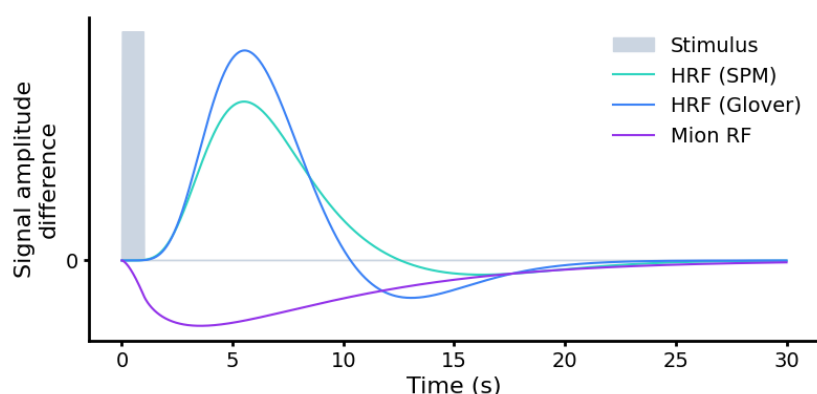


Figure 1.2: Examples of Hemodynamic and MION Response Functions Depending on the participant, on the area of the cortex, and on other factors, the shape of the HRF can vary greatly. We show two canonical examples of HRF coming from SPM and [Glover, 1999](#). We also show the response function for MION.

Adapted from the Nilearn documentation.

1.1.2 Functional MRI stimuli

Functional MRI seeks to study the dynamics of the brain, and significant efforts have investigated how these change when participants are involved in controlled activities.

1.1.2.1 Task-based protocols

A large proportion of fMRI studies involve participants performing specific tasks. These tasks can be as simple as watching images and indicating if they were presented before [[Haxby et al., 2001](#)], or as complex as solving mathematical problems [[Amalric and Dehaene, 2016](#)]. Stimuli are often organised in a so-called *block design*, which alternates between short periods of rest and task performance. Many difficulties arise when designing tasks for fMRI studies. In particular, researchers must design *control tasks* to isolate the brain's response to the task of interest. For instance, showing images of faces and non-face objects helps isolate areas of the brain that are more involved in face- than object-perception. In this case, viewing images of faces is the main task, and viewing images of non-face objects is the control task.

1.1.2.2 Task-free protocols

As task-based protocols may introduce biases in how researchers probe the brain's dynamics, more natural, task-free protocols have gained popularity.

Naturalistic stimuli In particular, so-called *naturalistic* stimuli are extensively used to study the brain's dynamics. These stimuli are designed to mimic the complexity of the real world and are thought to elicit more natural responses

from participants. [Nastase et al., 2020](#) compiles a series of arguments in favour of using naturalistic stimuli in fMRI studies if one wants to build generalisable and robust models of the brain. Naturalistic tasks range from passive tasks like watching short video clips [[Wen et al., 2017](#)] or listening to stories [[LeBel et al., 2023](#)], to more interactive tasks like playing video games [[Zhang et al., 2020](#)].

Resting-state Extensive research has been conducted on data collected from participants not involved in particular instructed tasks. In particular, the *resting-state* protocol scans immobile participants in the hope of capturing the brain's intrinsic dynamics. Interestingly, the authors of [Horikawa et al., 2013](#) have also acquired images in sleeping participants, seeking to reconstruct their visual perceptions while dreaming.

1.1.3 Analysing functional MRI data

Images collected with fMRI are noisy. Moreover, changes in blood oxygenation measured by BOLD imaging are of the order of a few percents and cannot be precisely detected by the naked eye. Therefore, researchers have consistently been developing statistical methods to analyse fMRI data, extract meaningful information from it, and link it with the properties of stimuli presented during the acquisition. We divide these methods into two categories: encoding and decoding models.

1.1.3.1 Encoding

Encoding models seek to predict the brain's response to stimuli based on the properties of these stimuli. While many approaches are possible, most encoding models are framed as linear optimisation problems in which the brain responses are modelled as a linear combination of the properties of the stimuli.

GLM The most commonly used encoding approach is the General Linear Model (GLM). It first derives a *design matrix* $\mathbf{X} \in \mathbb{R}^{t,r}$ which compiles information regarding the stimulus presented to the participant at each time point t , which we seek to correlate with the brain's activity. On top of these stimulus-related regressors, the design matrix also includes confounding regressors that could explain parts of the signal without being related to brain activity. [Figure 1.3](#) shows an example of a design matrix for a block-design experiment. These confounding factors mainly include the participant's head motion, or drifting effects in the scanner. In a block-design experiment, the design matrix would typically contain a column for each type of task block, with a 1 when the block is active and a 0 otherwise. These rows need to be convolved with an HRF to account for the delay between the presentation of the stimulus and the brain's response. Finally, we seek to derive $\beta \in \mathbb{R}^{r,n}$, the matrix of beta coefficients, which quantifies the relationship between each of the r types of stimuli and the activity $\mathbf{Y} \in \mathbb{R}^{t,n}$ of each of the n voxels in the brain:

$$Y = X\beta + \varepsilon$$

where ε models noise. If noise is assumed to be independently identically distributed (i.i.d.), the equation can be solved using ordinary least squares. However, it is known that the noise in fMRI data is not i.i.d. and exhibits temporal and spatial correlations. Therefore, it is common practice to use autoregressive models instead [Woolrich et al., 2001].

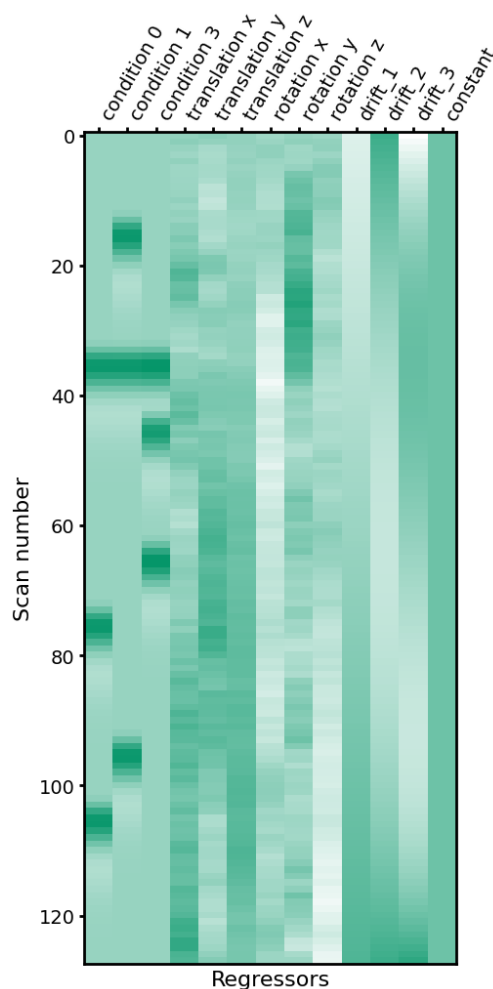


Figure 1.3: Example design matrix for a block-design fMRI experiment One seeks to determine areas of the cortex that are more active when the participant is presented with one type of stimulus – denoted as a condition – compared to another. Movements – translations and rotations – of the participant’s head are recorded during the acquisition and regressed out of the signal. Finally, a constant term and various drifts are added to the design matrix to account for the scanner’s drift.

Adapted from the Nilearn documentation.

Fixed-effect contrasts At this stage, the matrix of beta coefficients contains information linking areas of the brain with the stimuli presented to the partici-

pant. Still, it will not provide significant differences between stimuli in the general case. To do so, it is common practice to derive *contrasts* from the beta coefficients. These statistical tests seek to determine whether beta coefficients associated with a given voxel are significantly different when the participant is presented with one type of stimulus compared to another. Fitting a GLM and computing contrast maps completes the *first-level analysis* of fMRI data.

Second-level analysis The first-level analysis is typically conducted on each participant's data separately. In order to gain insights into the general population and statistical power, researchers often conduct a *second-level analysis* that aggregates results from all participants and subsequently runs a statistical test on these aggregated results.

1.1.3.2 Decoding

Conversely, decoding models seek to predict the properties of stimuli based on the brain's response to these stimuli. The seminal work of [Haxby et al., 2001](#) showed that it was possible to predict the category of objects presented to a participant based on fMRI data. They separated even and odd runs, and for each brain volume of the odd run, output the label of the most similar brain volume of the even run. [Figure 1.4](#) illustrates this nearest neighbour approach. It paved the way to *multi-voxel pattern analysis*. One caveat of this approach is that it requires researchers to identify the categories of objects they want to decode, which may introduce biases in the analysis. More recent work usually relies on pre-trained deep-learning models to get latent representations of shown stimuli, which partially alleviates this problem. These pieces of work usually fit a linear [[Ozcelik and VanRullen, 2023](#)] or non-linear [[Scotti et al., 2023](#)] decoding model to predict these latent representations from fMRI data. Predicted latent representations can then be used to infer the properties of stimuli presented to the participant or as inputs to pre-trained generative models to reconstruct the stimuli themselves [[Cheng et al., 2023](#)].

1.2 Functional brain mapping in human and non-human primates

Thanks to BOLD imaging and encoding models, researchers have studied the cortex's functional organization for the past 30 years. In particular, they have sought to identify areas of the cortex whose activity correlates with specific stimuli or tasks. Studying the functional organisation of the cortex is sometimes referred to as *functional brain mapping*. It tackles many fundamental questions in neuroscience, such as the existence of specialised areas, and the existence of similarities and differences in functional organisation between individuals as well as across species. This section presents some of the most critical findings in this field.

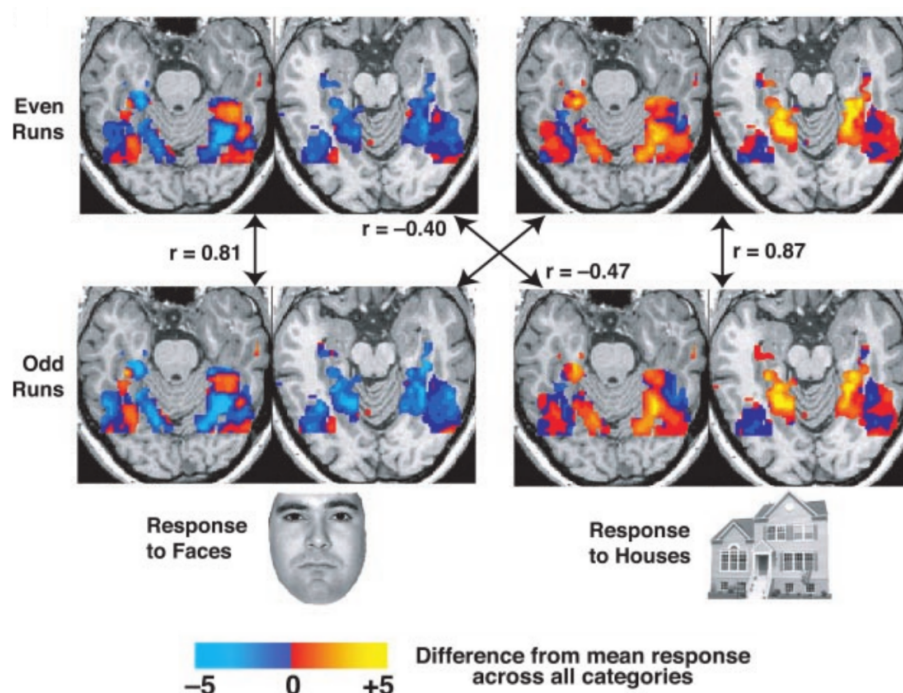


Figure 1.4: Decoding perceived stimuli by analysing distributed activations in the visual cortex Comparisons between the patterns of response to faces and houses in one subject. The within-category correlations for faces ($r = 0.81$) and houses ($r = 0.87$) are both markedly larger than the between-category correlations, yielding correct identifications of the category being viewed. Figure and caption from [Haxby et al., 2001](#).

1.2.1 Functional organisation of the cortex in humans

It is now common knowledge that the location of cortical areas involved in particular functions is not random, but follows general principles of spatial organisation, both in human and non-human subjects. There are strong pieces of evidence that particular bits of information are processed in specific parts of the cortex and that healthy individuals share a common global cortical organisation. In particular, primary areas – involved in motor, auditory, and visual capabilities – have been known to follow general principles of organisation for a long time. Indeed, the study of humans and animals who had survived different brain lesions showed that, in some cases, only certain aspects of their behaviour would be impaired by such lesions.

Historical landmarks Specifically, Marie-Jean-Pierre Flourens published a treatise in 1824 showing that the removal of specific parts of the cortex in pigeons and rabbits led to specific deficits in their behaviour [[Flourens, 1824](#)]. Similarly, in 1825, Jean-Baptiste Bouillaud compiled pieces of evidence showing that lesions in the frontal lobe in humans led to deficits in speech production [[Bouillaud, 1825](#)]. This research led Paul Broca to map the frontal cortex even more precisely and to discover the well-known Broca's area, which is involved

in speech production and linked to aphasic syndromes [Broca, 1865]. Moreover, armed conflicts caused neurologists to examine a large number of patients with brain lesions, which has allowed them to study the functional organisation of the occipital cortex in more detail. In particular, the study of soldiers who had suffered from gunshot wounds to the head during Russo-Japanese War by Tatsuji Inouye, and the First World War by Gordon Morgan Holmes, led to the discovery of the retinotopic organisation of the occipital cortex [Glickstein and Fahle, 2000; Holmes, 1917]. Retinotopy is the principle that the visual field is mapped onto the visual cortex in a systematic, diffeomorphic way, with neighbouring parts of the visual field being represented by neighbouring parts of the cortex.

Functional brain mapping with MRI Non-invasive imaging, and in particular fMRI, has allowed researchers to study the functional organisation of the cortex without the need to focus exclusively on individuals with brain lesions. At the end of the 20th and beginning of the 21st centuries, scientists started to use fMRI to study the functional organisation of the cortex in healthy individuals. For instance, Sereno et al., 1995 could push forward the work of Tatsuji and Holmes by establishing eccentricity and polar angle retinotopic maps of the visual cortex using MRI, as illustrated in Figure 1.5. In the following years, many research teams have used comparable imaging techniques to map the functional organisation of the visual cortex in humans, and discovered the existence of areas specialised in particular categories of stimuli.

It is the case of the Fusiform Face Area (FFA), which activates whenever a participant is shown an image of a face, with remarkable consistency across individuals [Kanwisher et al., 1997]; the Parahippocampal Place Area (PPA), which seems to specialise in visual stimuli showing places [Epstein et al., 1999]; the Extrastriate Body Area (EBA), which shows stronger activity for visual stimuli showing body parts [Downing et al., 2001]; the Visual Word Form Area (VWFA) which is anatomically close to the FFA and activates when participants are shown words [Cohen et al., 2000]. More recently, higher magnetic fields have allowed researchers to discover even more fine-grained mappings of these areas: for instance, the VWFA can be divided into sub-modules when readers are bilingual [Zhan et al., 2023]. Moreover, many more areas outside the occipital lobe have been shown to be significantly activated when performing specific, complex tasks. For instance, the Temporal Parietal Junction (TPJ) is involved in theory of mind tasks [Saxe and Kanwisher, 2013]; a fronto-parietal network was shown to be sensitive to the (un)predictability of stimuli and how much a participant is confident in their judgement [Bounmy et al., 2023].

Recent publications have aggregated a large number of publicly available fMRI datasets to provide a comprehensive view of the functional organisation of the human cortex [Mensch et al., 2021], as illustrated in Figure 1.6.

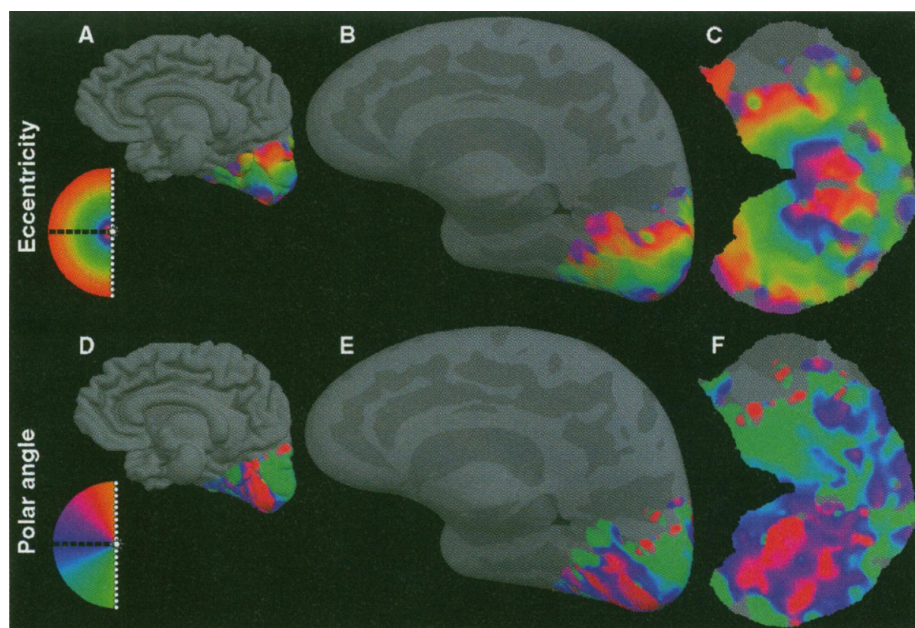


Figure 1.5: Isoeccentricity and isopolar angle maps of human visual areas. The top row shows isoeccentricity coded by color [red (fovea) → blue → green (parafoveal) → yellow → red (periphery)] displayed on the original cortical surface (A), the unfolded cortical surface (B), and the cut and flattened cortical surface (C). The bottom row shows polar angle [red (lower vertical meridian) → blue (horizontal meridian) → green (upper vertical meridian)] plotted on the same three surfaces (D), (E), and (F), respectively. Local eccentricity and polar angle were determined by considering the phase of the response to a slowly dilating ring or a slowly rotating hemifield at the dilation or rotation frequency. The unfolded representations in (B) and (E) were made by relaxing the curvature while approximately preserving local area and local angles (the sulcal cortex is dark gray and the gyral cortex light gray). The flattened representations in (C) and (F) were made with the same algorithm after the occipital lobe was cut off and an additional cut in the fundus of the calcarine sulcus was made.

Figure and caption from [Serenó et al., 1995](#).

Debates and questions Many questions arose from these discoveries. Firstly, some argued that not all cortical areas have a specialised function, and that some areas are involved in multiple tasks, like the multiple demand areas [[Duncan and Owen, 2000](#)] or the Default Mode Network (DMN). Nonetheless, these areas could still have a consistent location in the cortex across individuals.

Secondly, [Haxby et al., 2001](#) advocate that specific categories of visual stimuli can elicit activations in large parts of the cortex, which suggests that the specificity to certain categories of stimuli of cortical areas might not be all there is to understand about functional mapping. Moreover, they argue that some visual areas being more responsive to certain categories of objects does not imply that they are exclusively involved in processing these objects. In the particular case of language, authors of [Huth et al., 2016](#) show that listening to stories has been proven to activate a very large part of the cortex which can also be recruited for other tasks. It suggests that language processing is not limited to

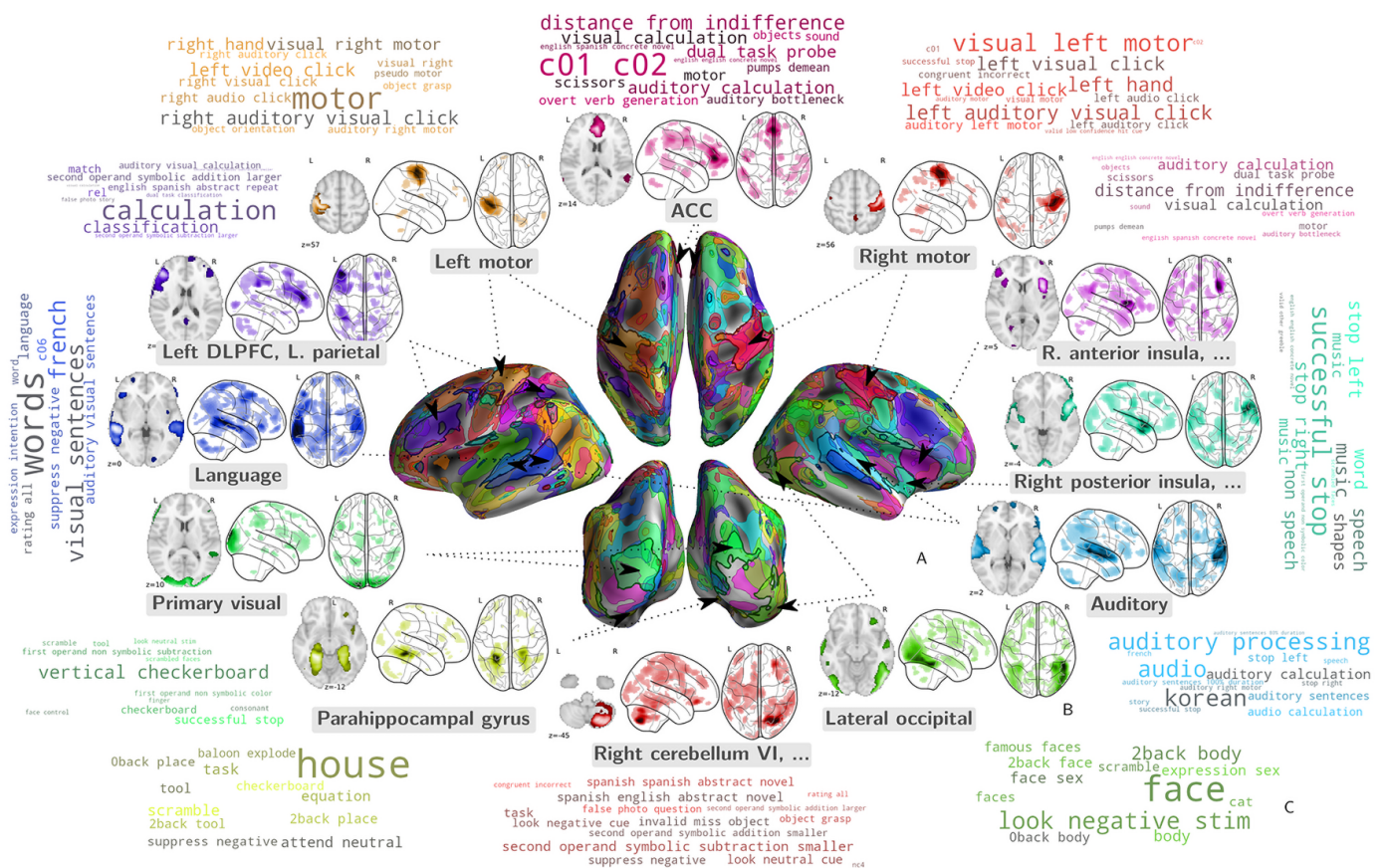


Figure 1.6: A cortical map of task-optimised networks The approach of [Mensch et al., 2021](#) learns networks that are important for decoding across studies. These networks are individually focal and collectively well spread across the cortex. They are readily associated with the cognitive tasks that they contribute to predict. We display a selection of these networks on the cortical surface (A) and in 2D transparency (B), named with the salient anatomical brain region they recruit, along with a word-cloud (C) representation of the stimuli whose likelihood increases with the network activation. The words in this word cloud are the terms used in the contrast names by the investigators; they are best interpreted in the context of the corresponding studies. Figure and caption from [Mensch et al., 2021](#).

a number of task-specific areas. Strikingly, the same study illustrates that these areas seemed to be selective of certain semantic categories, like places, time, or social interactions⁴.

Lastly, it is unclear whether these areas can also cause perception: if seeing a face triggers activity in the FFA, is it also true that activity in the FFA triggers the perception of a face? While the general question of causality is still open, [Schalk et al., 2017](#) showed that stimulating the FFA of one human participant with electrodes would cause them to report seeing faces – even though they were shown non-face objects – which suggests that the FFA could be causally involved in face perception.

⁴<https://gallantlab.org/viewer-huth-2016/>

1.2.2 Studying the functional organisation of other species to better understand the human brain

In the particular case of [Schalk et al., 2017](#), participants could only be implanted with electrodes because they were undergoing brain surgery as part of their treatment for epilepsy. This is not a common procedure, and it is generally not possible to stimulate the brain of healthy participants with electrodes. This is why researchers have been studying the functional organisation of the cortex in non-human primates – as well as other species – hoping that more invasive techniques would bring more insights. This approach is particularly relevant because the cortex of non-human primates shares similarities with that of humans.

The case of face-sensitive areas is of particular interest. While it was known from lesion as well as fMRI studies [[Tsao et al., 2003](#)] that the macaque's temporal lobe was involved in object recognition - and in particular face recognition - it was not clear whether the macaque's cortex had areas exclusively dedicated to face processing. In [Tsao et al., 2006](#), the authors used fMRI to find face-selective cortical regions in two macaques. Subsequently, they could probe the activity of single neurons in these areas using electrodes, and found that more than 90% of cells were very face-selective⁵ in the middle face patch – one of three identified face patches located in the temporal lobe.

In [Tsao et al., 2008](#), the authors could show that faces activate a network of areas in the macaque's cortex. Strikingly, they identified patches in the prefrontal cortex, one of which – termed as the Prefrontal Orbital (PO) – they show responds significantly more strongly to expressive macaque faces than to inexpressive faces. Other patches – including the Prefrontal Lateral (PL) and Prefrontal Arcuate (PA) – also showed increased activity when the macaques were shown expressive faces, but to a lesser extent. However, their location in the cortex made the authors hypothesise that they could be homologs of well-known areas in the human brain involved in language processing:

"The most posterior prefrontal face patch, PA, lies at the border between areas 44 and 45B. Area 44 is involved in fine control of facial musculature and may be the macaque homolog of Broca's area 13. The location of PA near area 44 may facilitate interaction between face perception and control of facial musculature, e.g., in mimicking behavior. [...] The strong right lateralization of PL raises the possibility that it may constitute the macaque homolog of the region in human right prefrontal cortex activated during remembrance of faces. It is then tempting to ask what is being coded by the corresponding left hemisphere region in the macaque – the answer may reveal a precursor of language." – excerpt from page 5 of [Tsao et al., 2008](#).

⁵In this study, the face-selectivity criterion implies that these cells show at least a 10:1 ratio of response to faces compared to other categories of objects

Interestingly, research in non-human primates is a stepping stone to building neuroprosthetics that could be used in impaired humans in the future. Much effort is put into decoding motor activity and intentions from the cortex. In their work, [Hamed et al., 2007](#) could decode the activity of the primary motor cortex of macaques implanted with electrodes reading from a few dozen neurons only. Their decoding accuracy was still very high for a given set of movements. Similarly, in [Astrand et al., 2014](#), the authors benchmarked a series of decoding algorithms on neural activity recorded from the prefrontal areas of macaques.

While it is yet unclear whether these pieces of information acquired in macaques are also valid to humans, it illustrates very well how knowledge acquired in one species can fuel research in another.

Comparison of functional organisation in visual areas Consequently, many research studies sought to directly describe the similarities and differences in the functional organisation of the cortex between human and non-human primates. Once again, a large proportion of them focused on the visual cortex. In particular, the review of [Orban et al., 2004](#) lists pairings of homologous areas in the lower and higher visual cortices of humans and macaques using fMRI. By definition, areas are considered homologous across species if they derive from the same area in a common ancestor. However, since one cannot study these now-extinct ancestors, researchers rely on similarities in anatomy, connectivity, cyto- and myelo-architecture, and function to infer homology. In this review, the authors list evidence showing that the retinotopic organisation of the early visual cortex – areas V1, V2, and V3 – is similar across humans and macaques. While area V3A seems to exist in both species, there is evidence that it is motion- and 2D-shape sensitive in humans but not in macaques. Similarly, the Middle Temporal (MT)+ complex in humans, involved in motion perception, seems homologous to the MT area in macaques. The authors also mention the similarities between the Inferior Temporal (IT) cortex of both species. However, a precise parcellation of this area is still lacking and might not be fully homologous across species.

Later, [Kriegeskorte et al., 2008](#) used fMRI to compare properties of the IT cortex, which is involved in categorising visual stimuli, in humans and macaques. They showed a series of objects to subjects of both species and computed, in each species, a matrix of dissimilarity between the responses of each pair of objects. They found that the dissimilarity matrices of both species were very similar, which suggests that the IT cortex encodes similar information across species. This study was one of the first to compare not only similarities in functional specialisation but also how different species perform this function.

Strikingly, the authors of [Kolster et al., 2009](#) showed that the middle temporal area – i.e., MT/V5 – and its neighbours are organised in clusters with a common foveal representation. This discovery was made possible by acquiring high-resolution fMRI data in macaques. In follow-up work, the authors of [Kolster et al., 2010](#) extended these results by showing that the potential counterpart func-

tional areas MT/V5+ in humans also exhibit a retinotopic organisation. Finally, the authors of [Mantini et al., 2012](#) argue that inter-species comparisons using prior knowledge of functional homologies can be detrimental to the discovery of better homologies. They propose to derive homologies between species using a data-driven approach based on the similarity of brain activity to rich stimuli. To this end, they compare the cortical activity of humans and macaques during a movie-watching task. They reveal that some functionally comparable areas now occupy very different locations in the cortex across species and that this reorganisation is not simply due to cortical expansion. Figure 1.7 illustrates this point.

The visual cortex has been extensively studied across both human and non-human primates for a series of pragmatic reasons: relative to the rest of the cortex, its volume is relatively large, making it easier to study with fMRI. Also, it is relatively easy to stimulate in highly controlled setups. However, other functional areas have been studied across species, with the striking example of [Sliwa and Freiwald, 2017](#). In this study, the authors show the existence of a ventrolateral prefrontal cortex area exclusively engaged in the analysis of social interactions, whose location could correspond to that of the Default Mode Network and Theory of Mind areas in humans.

Previously, the authors of [Joly et al., 2012](#) had studied the processing of vocalisations in humans and macaque, showing that the Superior Temporal Sulcus (STS) is activated in both species, with a preference for conspecific vocalisations.

Comparing network organisations The direct comparison of functional areas across species may not be all there is to understand about the functional organisation of the cortex. Consequently, many researchers have sought to compare networks of functional areas. In their seminal paper, the authors of [Neubert et al., 2014](#) use a combination of anatomical MRI and Diffusion Weighted (DW) MRI to derive a network of areas in the ventrolateral Frontal Cortex (vlFC), and resting-state fMRI to derive the functional connectivity profile of each of these areas. Finally, they compare the functional connectivity profiles of areas in the vlFC across the two species. The authors could identify 11 similar components in the vlFC of humans and macaques, and one component that was unique to humans.

Extensions to their work [[Mars et al., 2016, 2018](#)] prolong this comparison to the whole cortex, and use this data-driven approach to study the areal expansion of the human brain compared to that of macaques.

Interestingly, the authors of [Xu et al., 2020](#) adopt a comparable approach in which they seek to study whole-brain similarities and differences in cortical areas by relying on similarities in functional connectivity across a set of previously identified homologous cortical landmarks. They show that cross-species associations derived from these matchings can (1) predict other properties of the cortex, like patterns of myelin, or (2) be used to define a Functional Homol-

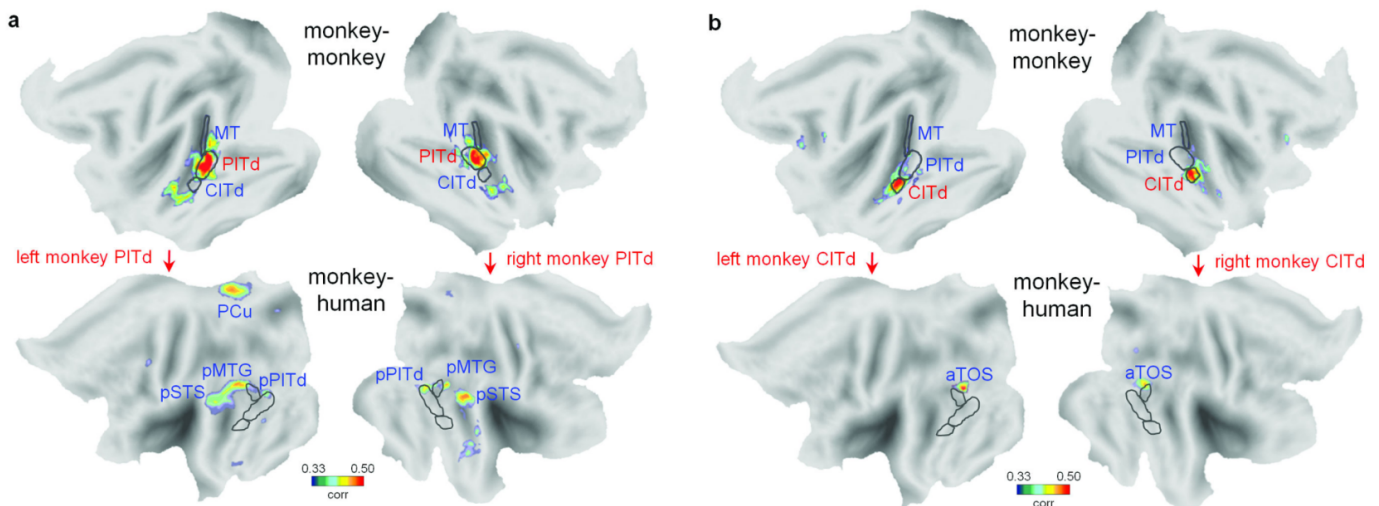


Figure 1.7: Intra- and inter-species activity correlation from monkey areas PITd and CITd Intra- and inter-species activity correlation maps (False Discovery Rate of $q < 0.001$) from both left and right monkey (a) PITd and (b) CITd. The correlation maps are shown only for the same hemisphere in which the seed area is positioned. The borders of monkey areas MT, PITd, CITd are drawn over the monkey flat map. The same borders after monkey-to-human cortical surface expansion are drawn over the human flat map. aTOS: anterior transverse occipital sulcus; PCu: precuneus; pMTG: posterior middle temporal gyrus; pPITd: human posterior area PITd. Figure and caption from [Mantini et al., 2012](#).

ogy Index quantifying the similarity of functional connectivity profiles across species in each area of the cortex, or also (3) provide insights about human cortical areal expansion.

Finally, the authors of [Eichert et al., 2020](#) use similar methods to study what differences between the cortical organisation of the human and macaque brain can be explained by areal expansion or tract extension only. In particular, they show that, contrary to other nearby cortical areas, the terminations of the arcuate fasciculus cannot be explained solely by these two factors, and thus conclude that it must have undergone additional evolutionary changes.

1.3 Inter-individual variability and related problems

The previous section showed that the functional organisation of the cortex and its comparison across species is a very active field of research. One could be tempted to think this problem could be solved by studying an arbitrarily large number of subjects in all species and by aggregating these results. This should in theory yield an arbitrarily precise map of the functional organisation of the cortex. However, this simple approach is not possible for one main reason: the anatomy and functional organisation of the cortex differ between individuals. Consequently, it is impossible to properly aggregate results from different subjects without the use of additional processing steps, referred to as *alignment* methods.

1.3.1 Anatomical alignment

The most commonly used method consists in registering each subject's brain images to a shared anatomy, referred to as a *template*. This way, each subject's brain is represented in the same space, and it is possible to compare the activity of a given cortical area across different individuals. This method is referred to as *co-registration* or *anatomical alignment*.

Many successful open-source pieces of software like FreeSurfer⁶ [Fischl, 2012], FSL⁷ [Jenkinson and Smith, 2001], ANTs⁸ [Avants et al., 2008], AFNI⁹ [Cox, 1996], or SPM¹⁰ [Friston et al., 1994] provide tools to compute these transformations. The registration process is usually one of many steps of a preprocessing pipeline. fMRIPrep¹¹ [Esteban et al., 2019] is a popular package that automates this pipeline and uses some of the tools mentioned above to compute the registration of fMRI data to a shared anatomy.

Let us distinguish between two types of alignment methods: those that align volumetric – or 3D – images, and those that align surfaces – or 2D – images. Volumetric alignment was the first to be developed and is still the most widely used. Initial alignment methods tried to maximise mutual information [Maes et al., 1997] or to implement variations of Demons algorithm [J.-P. Thirion, 1998]. Concurrently, the development of surface-based analysis methods in neuroscience – which can make visualisation and analysis simpler – led to the development of surface-based alignment methods. Notably, [Yeo et al., 2010] proposed a diffeomorphic method based on the spherical Demons algorithm. They use a spherical representation of the cortex and compute a deformation field that aligns the data of each individual to a template individual.

On top of an algorithm to compute the transformation, these methods also need a template anatomy. Consequently, a handful of such prototypical anatomies have been proposed. The Montreal Neurological Institute template MNI152 [Mazziotta et al., 1995], based on the brains of 152 participants, is the most widely used volumetric template in neuroimaging. The surface template associated with MNI152 is the `f3average` template [fischl_cortical_1999]. It is also widely used in surface-based analysis methods.

However, anatomical alignment makes two questionable assumptions: first, that anatomical information is sufficient to align individuals, and second, that the functional organisation of the cortex of aligned individuals is the same.

⁶<https://surfer.nmr.mgh.harvard.edu>

⁷<https://fsl.fmrib.ox.ac.uk>

⁸<http://stnava.github.io/ANTs>

⁹<https://afni.nimh.nih.gov/>

¹⁰<https://www.fil.ion.ucl.ac.uk/spm>

¹¹<https://fmriprep.org>

1.3.2 Variability in functional images

In the last decade, a growing body of research projects has resorted to scanning large numbers of participants. In particular, the Human Connectome Project (HCP) [Van Essen, Smith, Barch, Behrens, Yacoub and Ugurbil, 2013], the Adolescent Brain Cognitive Development (ABCD) [Casey et al., 2018], and the UK biobank [Sudlow et al., 2015] projects have scanned thousands of participants with mainly two goals: (1) to study the brain's functional organisation in more detail and (2) to link this imaging data with other phenotypical data, hoping that bio-markers involving brain imaging data could help the diagnosis of diseases. The initial intuition was that individual images are noisy, but the aggregation of many images would yield a more precise map of the brain's functional organisation.

Although initially modelled as noise, inter-individual variability in the functional organisation of the cortex is now recognised as a fundamental aspect of the brain, which the previously mentioned projects have both described and endured. In particular, it is a major impediment to defining atlases at the population level [Glasser et al., 2016]. Indeed, the authors of Glasser et al., 2016 show that the functional organisation of the cortex is highly variable across individuals, and that the functional areas of the cortex of one individual do not necessarily match those of another individual.

In fact, even areas considered to have a simpler topology and location across individuals may not be well characterised at the group level using standard techniques. For instance, the authors of Zhan et al., 2023 scanned about 20 participants with high-field fMRI. In particular, they acquire functional localisers to identify the exact position of the Fusiform Face Area (FFA) in each individual, and compare the location of the Visual Word Form Area (VFWA) and its sub-modules. Strikingly, they show that, without spatial smoothing, the FFA is not visible in group analysis at high spatial resolution, although it can clearly be identified in each participant. Applying spatial smoothing to the data allows them to identify the FFA at the group level, but this practice obviously blurs the exact location and specificity of cortical areas.

More generally, it is also not well understood what proportion of the signal acquired with fMRI is actually due to the task performed by the participant. In particular, the authors of Gratton et al., 2018 show that co-activity in functional networks seems to be driven by group- and individual-specific factors, and that the task performed by the participant only explains a small part of the signal.

Interestingly, within-subject variation should not always be modeled as noise, as neuroplasticity can lead to changes in the functional organisation of the cortex. For instance, Binda et al., 2018 use fMRI to study the effect of short-term deprivation visual cortex in adults. They show that even short periods of deprivation lead to significant differences in the BOLD signal; notably, it boosts V1 responses to the deprived eye. Castaldi et al., 2020 gives a more systematic

review of this type of studies.

Deep phenotyping Due to inter-individual variability and its effect on group statistics, many research teams have resorted to acquiring large amounts of data in a limited number of participants [Fedorenko, 2021]. Contrary to the previously mentioned large-scale projects that typically scan thousands of participants for approximately one hour each, these *deep phenotyping* projects typically scan a dozen of participants for tens of hours each. Many such datasets are openly available, like the Natural Scenes Dataset [Allen et al., 2022], the Courtois Neuromod Dataset ¹² or the Individual Brain Charting (IBC) dataset [Pinho et al., 2018]. The first two have acquired a very large number of samples of naturalistic stimuli, while the last one has reproduced many existing fMRI protocols hoping to map the functional organisation of the underlying participants precisely at the individual and group level. Figure 1.8 illustrates the concepts of inter-subject anatomical and functional variability, while giving a sense of the extent to which the brains of individual participants were mapped.

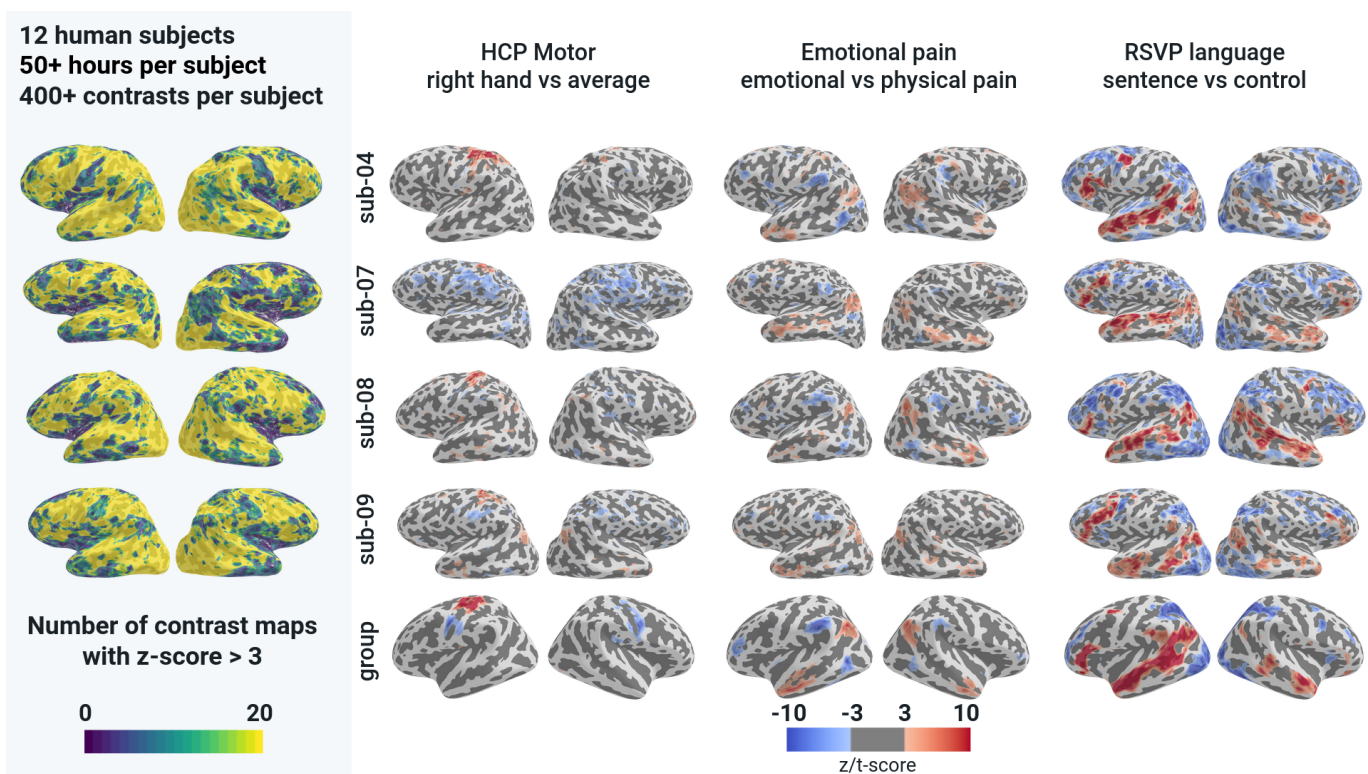


Figure 1.8: Inter-subject variability illustrated in the IBC dataset (Left) Participants of the IBC dataset have been extensively scanned with a wide range of tasks. Most of the cortex was significantly involved in at least twenty tasks in each subject. (Right) However, the anatomical and functional organisation of the cortex is highly variable across participants. Consequently, group-level statistics (bottom) do not always reflect the results observed at the individual level.

¹²<https://www.cneuromod.ca>

1.4 Towards functional alignment

In their seminal paper, [Sabuncu et al., 2010] propose to align cortices using functional data instead of relying solely on anatomical data. They use fMRI data acquired during a movie-watching task to compute alignments from each individual to a template individual. They show that this approach yields greater group statistics and generalises to other tasks. This approach is referred to as *functional alignment*. It is worth noting the authors introduce a surface-based alignment method, but that the concept of functional alignment can also be applied to volumetric data. In this study, the authors use a regularized mesh deformation model to change the position of vertices on the cortical sheet to (1) maximise the functional correlation between vertices while penalising (2) the appearance of folds in the warped mesh and (3) high differences in internode distances. They implement an iterative algorithm that updates the positions of vertices so that they can verify these three constraints. Moreover, they use a coarse-to-fine approach to prevent the algorithm from getting stuck in local minima.

Following this work, many other methods have been proposed to align the functional organisation of the cortex of different individuals. We divide them into two categories: diffeomorphic methods, which rely on a smooth transformation of the cortical mesh, and non-diffeomorphic methods, which do not make this assumption.

1.4.1 Existing voxel-to-voxel functional alignment methods

All methods presented in this section seek to derive voxel-to-voxel (or vertex-to-vertex) transformations. These transformations should match voxels (or vertices) from two different brains based on a similarity between their features.

1.4.1.1 Diffeomorphic methods

Multimodal Surface Matching (MSM) Among other methods relying on deformations of meshes, Multimodal Surface Matching (MSM) [Robinson et al., 2014] is a popular choice. It has been used in many studies to align the functional organisation of the cortex of different individuals, both in intra-species [Glasser et al., 2016] and inter-species setups [Eichert et al., 2020; Xu et al., 2020].

MSM implements a coarse-to-fine strategy to compute a deformation field that aligns the functional data of two individuals. At each step, each vertex v of the source mesh is associated with a *label* – which can be understood as a target vertex. This label is chosen from a discrete number of possible labels nearby such that this choice minimises a cost function that (1) estimates the overlap of cortical areal features (or shape) from the correlation of overlapping patches of features in source and target meshes (at the location corresponding to that proposed move) while (2) penalising choices of labels that would drag v away

from its neighbours. In their seminal work, the authors of [Robinson et al., 2014](#) formulated this regularisation term by penalising the geodesic distance on the mesh between each pair of moved vertices forming an edge of the mesh. This term was later reformulated to penalise areal changes of faces (rather than edges) in order to obtain smoother, more accurate solutions [[Robinson et al., 2018](#)].

In both cases, the optimisation problem is framed as a first order Markov Random Field problem. Many solvers exist for this framework, and the authors of [Robinson et al., 2014](#) use a pre-existing implementation of the FastPD algorithm, itself being based on the graph-cut algorithm [[Komodakis and Tziritas, 2007](#)].

1.4.1.2 Non-diffeomorphic methods

Other functional alignment methods that do not rely on a smooth transformation of the cortical mesh exist. Let us denote $\mathbf{F}^s \in \mathbb{R}^{n,k}$ and $\mathbf{F}^t \in \mathbb{R}^{p,k}$ the functional data of the source and target individuals, respectively. n and p are the number of vertices in the cortical mesh of the source and target individuals, respectively, and k is the number of features extracted from the fMRI data. The goal of these methods is to find an alignment $\mathbf{P} \in \mathbb{R}^{n,p}$ that aligns \mathbf{F}^s to \mathbf{F}^t .

Ridge regression Ridge regression gives a straightforward way to find a linear transformation \mathbf{P} that aligns \mathbf{F}^s to \mathbf{F}^t :

$$\mathbf{P} \triangleq \arg \min_{\mathbf{P}} \|\mathbf{P}^T \mathbf{F}^s - \mathbf{F}^t\|_2^2 + \lambda \|\mathbf{P}\|_2^2$$

where $\|\cdot\|_2$ denotes the L2 norm, and λ is a hyperparameter that controls the trade-off between the alignment of the data and the regularisation of \mathbf{P} . This simple method can serve as the go-to approach for functional alignment because it requires minimal tuning and is computationally efficient, although it can be used to tackle non-trivial tasks [[Ferrante, Boccato and Toschi, 2023](#)]. The work of [Bazeille, 2021](#) compares it to other methods extensively.

Hyperalignment In their seminal paper, [Haxby et al., 2011](#) propose a method called *hyperalignment* to align the functional data of different individuals. The transformation \mathbf{P} is constrained to be orthogonal, which ensures that the solution is unique and offers an interpretation of the transformation as a rotation in the space of voxels:

$$\mathbf{P} \triangleq \arg \min_{\substack{\mathbf{P} \\ \text{s.t. } \mathbf{P}^T \mathbf{P} = \mathbf{I}_p}} \|\mathbf{P}^T \mathbf{F}^s - \mathbf{F}^t\|_2^2$$

Initially, the authors designed this problem to find a rotation in the space of voxels that aligns two cloud points \mathbf{F}^s and \mathbf{F}^t between two individuals. These cloud points represent timeseries of brain activity in the space of voxels. Many

variations of hyperalignment have been proposed since then, as well as different evaluations of the method, and can be found in the review of [Haxby et al., 2020].

Optimal transport Comparably, Bazeille et al., 2019 proposes to derive linear transformations from an optimal transport formulation

$$P \triangleq \arg \min_{\substack{P \geq 0 \\ P_{\#1} = w^s \\ P_{\#2} = w^t}} \langle C, P \rangle = \arg \min_{\substack{P \geq 0 \\ P_{\#1} = w^s \\ P_{\#2} = w^t}} \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}} \|F_i^s - F_j^t\|_2^2 P_{i,j}$$

where $C \triangleq \left(\|F_i^s - F_j^t\|_2^2 \right)_{i,j} \in \mathbb{R}_+^{n,p}$ is the feature cost matrix, $w^s = \frac{1}{n} \mathbb{1}_n$ and $w^t = \frac{1}{p} \mathbb{1}_p$ are the distribution weights over the source and target meshes, and $P_{\#1} = (\sum_j P_{i,j})_i$ and $P_{\#2} = (\sum_i P_{i,j})_j$ denote the marginals of P .

We leave the details around the formulation of this problem to Chapter 2. However, let us already note that this method relaxes the orthogonality constraint of hyperalignment and allows for more expressive transformations.

1.4.2 Aligning multiple individuals with functional templates

Computing voxel-to-voxel transformations from two individuals does not allow one to leverage data collected in a large number of participants. Therefore, many adaptations to the methods presented previously have been developed to align the functional data of multiple individuals to a common space. Similarly to anatomical co-registration, functional alignment has sought to derive population templates that contain functional information derived from a group of individuals and that are hopefully more expressive than anatomical templates.

Shared Response Model In P-H. Chen et al., 2015, the authors introduce the Shared Response Model (SRM). Given functional data from m individuals, the authors propose to derive a shared response S that explains the functional data of multiple individuals, and a set of transformations P_1, \dots, P_m that project the functional data of each individual to S . We see it as a follow-up work from Haxby et al., 2011. As such, it implements the same constraint that estimated transformations should be rotations:

$$\min_{\substack{P_1, \dots, P_m, S \\ \text{s.t. } P_i^T P_i = I}} \sum_{i=1}^m \|F^i - P_i S\|_2^2$$

The authors implement a two-step algorithm to solve this problem: at each step, they alternate between computing the shared response S and the transformations P_1, \dots, P_m . Note that the following optimisation problem arises naturally, and is more similar to the ones presented in the previous paragraphs:

$$\begin{aligned} \min_{P_1, \dots, P_m, S} \quad & \sum_{i=1}^m \|P_i^T F^i - S\|_2^2 \\ \text{s.t. } \quad & P_i^T P_i = I \end{aligned}$$

However, in the original paper, the authors empirically find that the solutions computed for S are more meaningful with the first formulation than with the second. Many other methods have been proposed to build functional templates from multiple individuals. They usually formulate optimisation problems close to that of the SRM but with different constraints or regularisations. In particular, [Richard, Gresele, Hyvärinen et al., 2020](#) focuses on solving an Independent Component Analysis (ICA) problem, which replaces the orthogonality constraint with an invertibility constraint on P_1, \dots, P_m . They show that this allows them to derive more accurate transformations and templates on fMRI movie-watching data.

Projecting individual data onto a common latent space with deep learning

Because of inter-subject variability, recent deep learning approaches tackling the problem of training models that decode brain activity have often resorted to training a different model for each individual. However, some recent works have shown that it is possible to train a single model on data from multiple individuals and achieve good performance. In particular, [Défossez et al., 2023](#) adopt this approach on M/EEG data of participants listening to various auditory stimuli, and [Scotti et al., 2024](#) on fMRI data of participants watching images. These works have in common that the first layers of their model differ from participant to participant, but the last layers are shared across participants. These first individual layers can be understood as mappings of individual data to a shared latent space. While these approaches do not provide interpretable voxel-to-voxel transformations, nor stem from an explicit optimisation problem, they empirically lead to the construction of valuable shared functional spaces.

1.4.3 A critical perspective on current functional alignment methods

Alignment methods presented in the previous sections come with advantages and drawbacks. Let us focus on one crucial aspect that limits their use in practice: anatomical consistency. Indeed, users of functional alignment techniques usually expect that voxels / vertices that are matched together across individuals are located in comparable areas of the cortex. However, this constraint is hard to implement and potentially limits the expressiveness of the transformations that can be derived.

In that sense, diffeomorphic methods like MSM are the most conservative ones. They ensure that the cortical mesh of each individual is not deformed too much, which guarantees that the matched voxels are located in comparable areas of the cortex. Subsequently, computed alignments are interpretable and display solid anatomical consistency. However, they restrict alignments to a set

of transformations that cannot express all the variability in the functional data of different individuals. Indeed, the assumption that there exists a continuous mapping between the cortical surfaces of humans and macaques seems too strong [Eichert et al., 2020]. Moreover, this assumption is also questionable in the case of humans, as Glasser et al., 2016 shows that areas of the prefrontal cortex can vary widely in both location and topology across non-pathological human participants.

On the other hand, non-diffeomorphic methods like hyper-alignment can match voxels located in arbitrary areas of the cortex. For instance, nothing prevents these techniques from matching vertices in the visual cortex to the frontal lobe due to long range functional correlations – however this does not make sense based on what is known about cortical cyto-architecture in these cortical regions. In practice, they such matchings do happen and question the reliability of the underlying methods. In order to limit this drawback, non-diffeomorphic alignment methods usually restrict the set of possible solutions so that they are more anatomically consistent. In practice, this can be done by using a pre-computed atlas and computing a specific alignment for each underlying Region of Interest (ROI). However, this approach somewhat defeats the purpose of using whole-brain alignment methods as they rely on prior alignment assumptions. Moreover, this framework will generally exhibit artifacts at the borders of these regions.

In order to address these artifacts, researchers often resort to using a *searchlight* approach [Guntupalli et al., 2016] in which the brain is divided in a series of overlapping spheres. One alignment is computed for each sphere, and the results are then aggregated. The main limitations of this approach is that it once again relies on prior anatomical alignment assumptions, and that the aggregated alignments are no longer guaranteed to respect initial optimisation constraints like orthogonality or invertibility.

In conclusion, there is room for an alignment method which would provide anatomical consistency similar to that of MSM while relaxing the diffeomorphic constraint to enhance expressivity. Moreover, non-diffeomorphic methods also rely on searchlight or ROI-based approaches for computational reasons – because voxel-to-voxel transformations in high resolution volumes or meshes have a very large memory footprint – and it would be interesting to see if they can be extended in whole-brain setups while being freed from these constraints.

1.4.4 The need for open datasets, implementations and benchmarks

To this day, it is unclear which of these methods works best [Bazeille et al., 2021; Ho et al., 2023]. Indeed, while there is genuine interest from the neuroscience community to develop these techniques, comparing them is still very challenging. Consequently, choosing the most adapted method for a given task is not straightforward. Thus, these methods have seldom been used in research studies and are still very far from being used routinely.

We advocate that there are many important reasons for this situation. First, this field generally lacks identified tasks that could be used to compare these methods. As exposed in the previous sections, deriving voxel-to-voxel transformations in an inter-subject setup is not the only task of interest; alignment between tasks in an intra-subject setup is also an active topic of research, as well as the construction of templates.

Secondly, public datasets associated with these tasks do not always exist or are not clearly identified. Moreover, standard baselines and evaluation metrics are still scarce.

Lastly, finding simple, open-source implementations allowing one to use these methods, access these datasets, and run these benchmarks is still challenging.

Chapter 2

Introduction to Optimal Transport

2.1 A naive approach to functional alignment

It is natural to think that the problem of functional alignment can be solved by simply matching the vertices of the source and target meshes that are functionally closest to each other. However, this naive approach is not sufficient to ensure a meaningful alignment. In this section, we will briefly illustrate why.

We will consider a simple experiment. We take two individuals from the IBC dataset and take 370 z-scored contrast maps computed using task-based experiments that both participants performed. Let us denote the first participant as the source subject and the second as the target subject. For each vertex of the source, we will find the closest vertex of the target in terms of Euclidean distance in the 370-dimensional functional space.

Figure 2.1 shows that this method does not yield a meaningful alignment. Indeed, the number of target vertices matched to at least one source vertex is very low, with a few vertices always being chosen. However, we have seen in the previous Chapter that IBC participants have been extensively scanned, and that meaningful functional data should be available for a majority of cortical areas, so we would expect most vertices in the target subject to have a meaningful counterpart in the source subject.

Moreover, we show that this negative result holds when one normalises or standardises the functional data of the source and target subjects along the vertices or features dimension.

Figure 2.2 shows the first two principal components of the functional data of the concatenated source and target subjects. Although the source and target distributions overlap in some areas of the vector space, one can see that vertices of the target brain that accumulate the highest number of matched source ver-

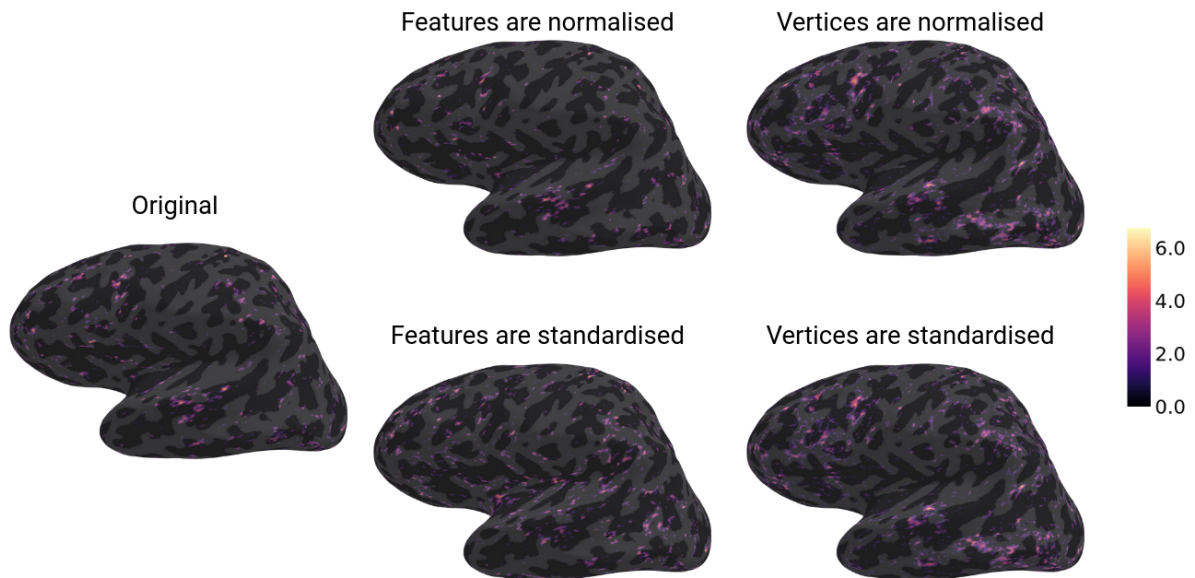


Figure 2.1: Count of vertices matched in the target brain For each vertex of the target brain, we show $\log(c + 1)$ where c is the number of vertices of the source brain that are matched to it. No matter the normalisation or standardisation applied to the functional data, the number of vertices in the target brain that obtain at least one match is low. In contrast, a small number of target vertices are matched up to $e^7 \approx 1000$ times.

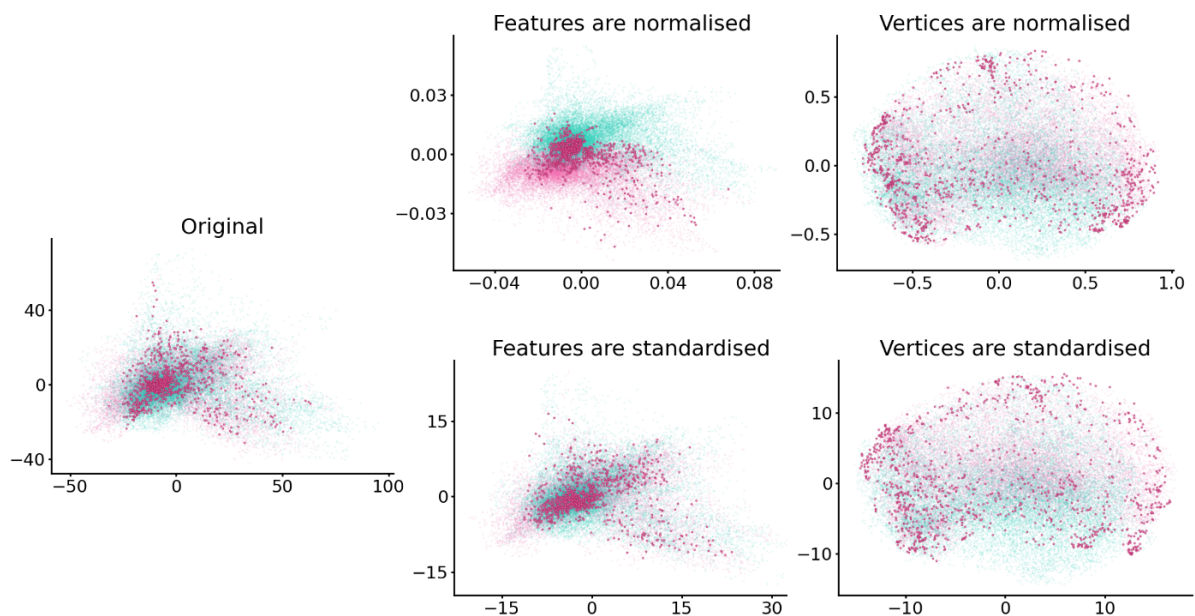


Figure 2.2: Principal components of the functional data across subjects Given 370 z-scored contrast maps computed using task-based experiments in two participants (pink and green) of the IBC dataset, we show the first two principal components in 20k (out of 150k) vertices in five previously mentioned setups. The top 1k (out of 150k) vertices of the target subject with the highest count are displayed in bold pink. They accumulate more than 50% of the source vertices.

tices do not necessarily span across the entire cloud points of the source and target individuals.

This simple experiment shows that it is not trivial to ensure that a large number of source and target vertices are matched. In the following sections, we will introduce the theory of optimal transport and show how it naturally introduces such incentives in the form of *marginal constraints*.

2.2 Principles of Optimal Transport for Machine Learning

2.2.1 Historical perspective

2.2.1.1 The Mines and Factories problem

The Mines and Factories problem was introduced by Leonid Kantorovich in 1942 [Kantorovich, 1942]. At the time, the Soviet Union was actively involved in World War II, and the government was looking for ways to optimize the allocation of resources to the front lines. However, the problem of finding the optimal way to transport resources from many mines to many factories is not easily solved, and sub-optimal solutions can be very costly. Kantorovich, who was working on optimally allocating resources for the Soviet army, formulated the optimisation problem presented below.

Let $\mathbf{a} \triangleq (a_1, \dots, a_n)$ be the vector representing the quantity of supply at each of the n mines, and $\mathbf{b} \triangleq (b_1, \dots, b_p)$ be the vector of demands at each of the p the factories. We assume that \mathbf{a} and \mathbf{b} are normalised, i.e. $\sum_i a_i = \sum_j b_j = 1$. Let $\mathbf{C} \in \mathbb{R}^{n \times p}$ be the matrix of costs, where $C_{i,j}$ is the cost of transporting one unit of supply from mine i to factory j .

The so-called *Monge-Kantorovich problem* is to find $\mathbf{P} \in \mathbb{R}^{n \times p}$, where $p_{i,j}$ is the quantity of supply transported from mine i to factory j , such that all goods are transported and the total cost of transportation is minimal:

$$\begin{aligned} \min_{\mathbf{P} \in \mathbb{R}_+^{n,p}} \langle \mathbf{C}, \mathbf{P} \rangle &\triangleq \min_{\mathbf{P} \in \mathbb{R}_+^{n,p}} \sum_{\substack{0 \leq i < n \\ 0 \leq j < p}} C_{i,j} P_{i,j} & (2.1) \\ \text{s.t. } P_{\#1} &= \mathbf{a}, \\ P_{\#2} &= \mathbf{b} \end{aligned}$$

We denote $P_{\#1} \triangleq \mathbf{P} \cdot \mathbb{1}_p$ and $P_{\#2} \triangleq \mathbf{P}^T \cdot \mathbb{1}_n$ the first and second marginal of \mathbf{P} . Here, the condition $P_{\#1} = \mathbf{a}$ ensures that, for each mine, the total quantity of supply transported is equal to the quantity of supply available in this mine. Respectively, $P_{\#2} = \mathbf{b}$ ensures that, for each factory, the total quantity of supply received is equal to the quantity of supply demanded by this factory.

2.2.1.2 Historical landmarks and recent developments

Kantorovich's optimisation problem is a generalisation of the *assignment problem* introduced by Gaspard Monge at the end of the 18th century – and thus also

known as the *Monge problem* [Monge, 1781]. At the time, Monge only considered the case where $n = p$ and P is a permutation matrix, thus forcing the supply from each mine to be transported to a single factory. Kantorovich’s generalisation, known as the *Kantorovich relaxation*, allows for a more flexible allocation of resources, where the supply from each mine can be split and transported to multiple factories.

Tolstoi and Hitchcock independently formulated the Mines and Factories problem, and the latter proposed a method to solve it in 1941 [Hitchcock, 1941; Tolstoi, 1930]. George Dantzig applied the simplex method to solve the Monge-Kantorovich problem in 1951 [Dantzig, 1951], opening the door to the use of linear programming for solving transportation problems. Kantorovich shared the 1975 Nobel Prize in Economics with Tjalling Koopmans for their work on the theory of optimal allocation of resources.

Later theoretical works from Brenier and Villani have further rooted the theory of optimal transport in the field of mathematics [Brenier, 1991; Villani, 2003]. More recently, Cuturi showed that the optimal transport problem can be solved efficiently using the Sinkhorn-Knopp algorithm [Cuturi, 2013], paving the way for GPU-accelerated optimal transport solvers. Since then, the field has seen a surge of interest in the machine learning community, where optimal transport is used for tasks such as generative modelling and domain adaptation [Arjovsky et al., 2017; Courty et al., 2016].

2.2.2 Framing brain functional alignment as an Optimal Transport problem

As seen in Section 2.1, naive approaches to functional alignment lead to problematic alignments. In particular, we showed that not all brain regions of a target brain are reached by these naive alignments. Consequently, it is natural to formulate an optimisation problem that includes constraints similar to that of the Mines and Factories problem.

Given a source mesh of size n , we associate each vertex of the mesh with a weight – or *mass* – $\frac{1}{n}$ and denote as $\mathbf{a} \in \mathbb{R}^n$ the vector of weights. Respectively, we associate each vertex of the target mesh with a weight $\frac{1}{p}$ and denote as $\mathbf{b} \in \mathbb{R}^p$ the vector of weights. We want to *transport* the mass from each vertex of the source mesh to one or multiple vertices of the target mesh, based on the similarity between the signals associated with the vertices. Therefore, we define a cost matrix $\mathbf{C} \in \mathbb{R}^{n,p}$, where $C_{i,j}$ denotes the similarity between the signals associated with the i -th vertex of the source mesh and the j -th vertex of the target mesh. Finally, we frame our brain alignment problem as the optimisation problem presented in Equation 2.1.

Note that not all vertices need to have the same mass, but without additional information on the respective sizes of the brain regions in the source and target brains, we assign the same mass to all vertices. Besides, in the following

chapters, we will denote as w^s and w^t the vectors of weights associated with the vertices of the source and target meshes, respectively.

2.2.3 Different classes of optimal transport problems

OT is rich with a lot of theory related to general measures – as opposed to discrete measures. However, in this thesis, we are only interested in applications of OT to machine learning. In this context, we only focus on representations of information that machines can process, i.e., vectors, matrices or more generally tensors. These are often sampled from analogous signals. Therefore, we will only consider formulations of OT problems on discrete measures.

We refer the interested reader to [Peyré and Cuturi, 2020](#) for a more in-depth introduction of the theory of OT, both from a theoretical and applied point of view.

In the following paragraphs, we will introduce different classes of OT problems, which will be used throughout this thesis. We will do so by using discrete formulations.

2.2.3.1 Balanced and unbalanced problems

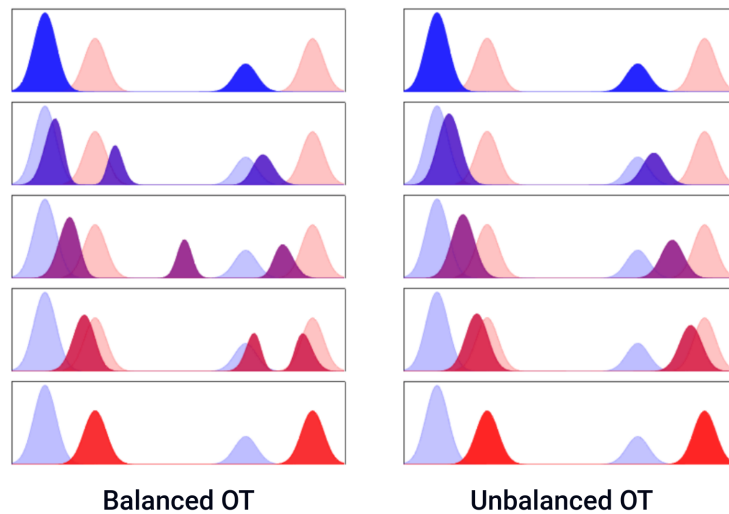


Figure 2.3: Balanced vs unbalanced optimal transport problems (Left) In the balanced case, if one tries to transport the blue and red distributions, some mass from the left mode of the blue distribution will have to be transported to the right mode of the red distribution. (Right) In the unbalanced case, with the proper hyper-parameter values, none of the left blue mode will be transported to the right red mode. Figure adapted from [[Peyré and Cuturi, 2020](#)], section 10.2.

Equation 2.1 is an example of *balanced* OT problem: it forces the total quantity of supply transported from the mines to be equal to the total quantity of supply demanded by the factories. Relaxing this constraint leads to *unbalanced* OT problems, with the following formulation:

$$\min_{\mathbf{P} \in \mathbb{R}_+^{n,p}} \langle \mathbf{C}, \mathbf{P} \rangle + \rho_1 d(\mathbf{P}_{\#1} \mid \mathbf{a}) + \rho_2 d(\mathbf{P}_{\#2} \mid \mathbf{b}) \quad (2.2)$$

where $\rho_1, \rho_2 \in \mathbb{R}_+$ are the unbalancing factors and d is a distance between two vectors (in the general case, d should be a divergence between two measures, but, since we only consider discrete measures, we will use the term distance for simplicity).

In this setup, higher values of ρ_1 and ρ_2 will force the marginals of \mathbf{P} to be close to \mathbf{a} and \mathbf{b} , respectively. Lower values will allow for more flexibility. The two new terms are referred to as *marginal constraints*.

They come with an intuitive interpretation in the context of brain functional alignment. Let us consider the following situation: it can very well be that a given functional area is bigger – in the sense that it covers more vertices – in the source brain than in the target brain. In this case, the ideal transport plan should *shrink* the source area to match the target area. It would result in the target area receiving more mass than it should accept, violating the initial balanced constraint. Figure 2.3 illustrates this situation.

Relaxing this hard constraint allows to model changes in the size of functional areas between the source and target brains, and to formulate it as a trade-off between being matched to functionally different areas and being more / less transported. An extreme case may arise when comparing cortical surfaces from different species, where some functional areas may exist in one species and not in the other.

In that sense, the unbalancing factor ρ_1 can be seen as weighting the importance of transporting brain regions of the source brain even though they cannot be functionally matched with areas of the target brain. Therefore, the value of ρ_1 should be chosen so that the functional cost of aligning two functionally different regions is higher than the cost of not aligning them. However, this sweet spot is not easy to find.

An analogous interpretation can be made for ρ_2 . In practice, we will usually set $\rho_1 = \rho_2$, since the source and target meshes have approximately the same amount of vertices.

2.2.3.2 Wasserstein and Gromov-Wasserstein problems

Equation 2.1 is an example of a *Wasserstein* OT problem: it assumes that the source and target functional data belong to the same vector space and can therefore be compared directly.

In other words, given $\mathbf{F}^s \in \mathbb{R}^{n,d}$ and $\mathbf{F}^t \in \mathbb{R}^{m,d}$, the d -dimensional vectors representing the functional signals associated with the vertices of the source and target subjects respectively, we assume that the quantity $\|\mathbf{F}_i^s - \mathbf{F}_j^t\|_2$ is meaningful for all i and j , and thus define the cost matrix $\mathbf{C} \triangleq \left(\|\mathbf{F}_i^s - \mathbf{F}_j^t\|_2^2 \right)_{i,j}$.

However, it may be the case that the functional data of the source and target subjects belong to different vector spaces and, therefore, cannot be compared

directly across individuals. OT can still be used in this case by defining a cost matrix G that takes into account the differences between the source and target functional data, as exemplified in Figure 2.4.

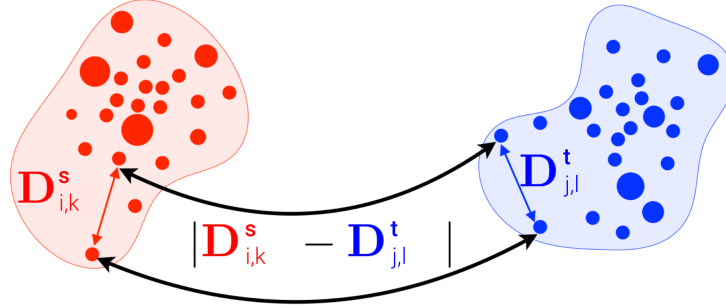


Figure 2.4: Gromov-Wasserstein problems Gromov-Wasserstein formulations aim to preserve the inner structure of the source and target measures, although they may belong to different spaces. Figure adapted from [Peyré and Cuturi, 2020], section 10.6.3.

This leads to the *Gromov-Wasserstein* problem [Mémoli, 2011], which is defined as follows:

$$\begin{aligned} \min_{P \in \mathbb{R}_+^{n,p}} \langle G, P \otimes P \rangle &\triangleq \min_{P \in \mathbb{R}_+^{n,p}} \sum_{\substack{0 \leq i,k < m \\ 0 \leq j,l < n}} G_{i,j,k,l} P_{i,j} P_{k,l} & (2.3) \\ \text{s.t. } P_{\#1} &= a, & \text{s.t. } P_{\#1} &= a, \\ P_{\#2} &= b & P_{\#2} &= b \end{aligned}$$

where

$$\begin{aligned} G &\triangleq \left(\left| D_{i,k}^s - D_{j,l}^t \right|^2 \right)_{i,j,k,l} &\in \mathbb{R}^{n,m,n,m} \\ D^s &\triangleq \left(\left\| F_i^s - F_k^s \right\|_2^2 \right)_{i,k} &\in \mathbb{R}^{n,n} \\ D^t &\triangleq \left(\left\| F_j^t - F_l^t \right\|_2^2 \right)_{j,l} &\in \mathbb{R}^{m,m} \end{aligned}$$

Here, G is the cost matrix, D^s and D^t are the distance matrices within the source and target functional data, respectively. The Gromov-Wasserstein problem seeks to match vertices together such that, if i and k are far apart in the source brain, they should be matched respectively to j and l , which are also far apart in the target brain.

Contrary to the Wasserstein problem, the Gromov-Wasserstein problem involves matching individual points while considering couplings between entire metric spaces. Moreover, it is a non-convex problem. Thus, it is generally harder to solve than the Wasserstein problem.

Let us mention two types of brain data that cannot be compared using a Wasserstein loss, but that fit the Gromov-Wasserstein framework.

First, resting-state fMRI data typically falls in this category. Indeed, let us consider $F^s \in \mathbb{R}^{n,t}$ and $F^t \in \mathbb{R}^{m,u}$ two individual timeseries of resting-state acquired in two participants. Nothing guarantees that the timeseries from the

two participants have the same length, and therefore they cannot be compared directly. Even in the case where $t = u$, nothing guarantees that the cognitive experience in one participant is synchronised with that of the other participant. However, comparisons of timeseries within the same participant are meaningful. In fact, researchers have heavily relied on this assumption to define within-subject functional connectivity matrices, which they would subsequently compare across participants [Finn et al., 2015]. However, this requires defining a shared set of comparable ROI across individuals, which we advocate defeats the purpose of using these inherently idiosyncratic datasets. However, the Gromov-Wasserstein loss yields a natural framework to compare this data without relying on those priors.

Secondly, un-registered anatomical data also falls in this category. Indeed, one cannot properly define the anatomical distance between vertices of two different brains. However, the geodesic distance from a vertex to all other vertices within the same brain is meaningful. The Gromov-Wasserstein loss allows one to rely on these well-defined distances to derive alignments.

Fused Gromov-Wasserstein problem Interestingly, it is possible to design optimisation problems using both the Wasserstein loss and Gromov-Wasserstein loss. These are hence called *Fused Gromov-Wasserstein* problems.

$$\begin{aligned} \min_{\mathbf{P} \in \mathbb{R}_+^{n,p}} \quad & \alpha \langle \mathbf{C}, \mathbf{P} \rangle + (1 - \alpha) \langle \mathbf{G}, \mathbf{P} \otimes \mathbf{P} \rangle \\ \text{s.t.} \quad & \mathbf{P}_{\#1} = \mathbf{a}, \\ & \mathbf{P}_{\#2} = \mathbf{b} \end{aligned} \quad (2.4)$$

where $\alpha \in [0, 1]$ is a hyperparameter that balances the relative importance of the Wasserstein loss and the Gromov-Wasserstein loss.

2.2.3.3 Unregularised and regularised problems

In the general case, solving for Equation 2.1 is computationally expensive, and has a minimal complexity of $O(n^3 \log(n))$ where n is the dimension of compared discrete measures [Pele and Werman, 2009]. A natural idea is to add a regularisation term to the problem, making the optimisation problem easier to solve [Wilson, 1969]. The entropy function $H(\mathbf{P}) \triangleq - \sum_{i,j} \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j})$ is often used as a regularisation term, leading to the following formulation:

$$\begin{aligned} \min_{\mathbf{P} \in \mathbb{R}_+^{n,p}} \quad & \alpha \langle \mathbf{C}, \mathbf{P} \rangle - \varepsilon H(\mathbf{P}) \\ \text{s.t.} \quad & \mathbf{P}_{\#1} = \mathbf{a}, \\ & \mathbf{P}_{\#2} = \mathbf{b} \end{aligned} \quad (2.5)$$

The intuition behind this addition is that solutions to the *unregularised* problem of Equation 2.1 are located on vertices of the simplex. Classical Linear Program-

ming approaches will explore all vertices of the simplex, which can be computationally expensive. Adding a regularisation term will push the solution away from the simplex's boundaries and towards its interior, making them easier to track. Moreover, this regularisation turns a convex problem into a strongly convex problem, which guarantees the uniqueness of the solution. Note that the transport plan computed with this regularised problem will be more *dense* than the one computed with the unregularised problem, as illustrated in Figure 2.5. Indeed, a transport plan P that matches all source and target vertices together with equal probability will be very dense, while the entropy $H(P)$ will be null. Moving away from a transport plan which matches vertices with equal, uniform probability will add information and make the quantity $-H(P)$ increase, at a rate proportional to ε .

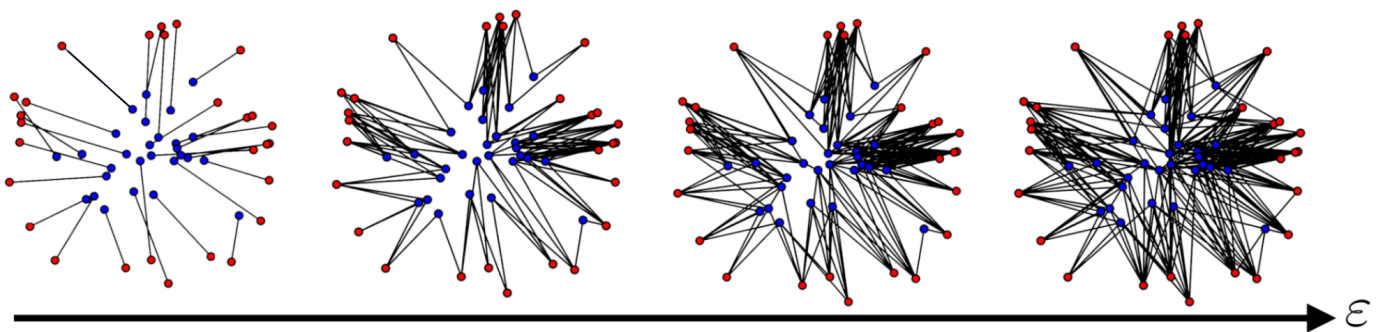


Figure 2.5: Effect of entropic regularisation on computed plan The hyper-parameter ε controls the amount of regularisation applied to the optimal transport problem. When ε is small, the optimal transport plan matching red and blue nodes is more sparse, and when ε is large, the optimal transport plan is more dense. Figure adapted from [Peyré and Cuturi, 2020], section 4.1.

Finally, a great advantage of using *entropic regularisation* is that its solutions approximate unregularised solutions, with a complexity of $O(n^2)$ [Cuturi, 2013]. This stems from the fact that solutions to Equation 2.5 can be shown to be of the particular form $P = UKV$, where $K \triangleq \exp(-C/\varepsilon)$, $U \triangleq \text{diag}(u)$, $V \triangleq \text{diag}(v)$, and u and v must verify $UKV \mathbb{1}_p = a$ and $VK^T U \mathbb{1}_n = b$. The classical proof consists in deriving the problem's Lagrangian: at optimality, the first-order derivative of the Lagrangian with regard to P is null, yielding the previous equality for P . The constraints on u and v emerge from the marginal constraints.

The problem of finding (u, v) in the previous setup is well-known, and the Sinkhorn theorem states that, for a given matrix K of positive elements, there exists a unique (u, v) – up to a multiplicative constant – such that UKV is doubly stochastic – i.e. the rows and columns sum to $\frac{1}{n} \mathbb{1}_n$ and $\frac{1}{p} \mathbb{1}_p$, respectively.

The Sinkhorn-Knopp algorithm provides an efficient way to compute u and v : it iteratively updates u and v such that the rows and columns of UKV sum to $\mathbb{1}_n$ and $\mathbb{1}_p$, respectively. We can very easily adapt this to the case where the rows

and columns should sum to a and b , respectively, which is the case in our OT problem. Algorithm 2.1 gives an explicit implementation of the Sinkhorn-Knopp algorithm for solving the Wasserstein problem with entropic regularisation.

Algorithm 2.1 Sinkhorn-Knopp algorithm for the entropic Wasserstein problem

Input: C, a, b, ε

Output: Optimal coupling P

- 1: $K \leftarrow \exp(-C/\varepsilon)$
 - 2: $u \leftarrow \mathbb{0}_n$
 - 3: $v \leftarrow \mathbb{0}_p$
 - 4: **while** (u, v) has not converged **do**
 - 5: $u \leftarrow a \oslash (Kv)$
 - 6: $v \leftarrow b \oslash (K^T u)$
 - 7: **end while**
 - 8: $P \leftarrow \exp(u \oplus v - C/\varepsilon)$
-

2.2.3.4 Combinations of Optimal Transport problems

Interestingly, it is possible to combine the different classes of OT problems presented above. For instance, regularised unbalanced Wasserstein problems can be solved efficiently using a slight adaptation of the Sinkhorn-Knopp algorithm. Moreover, it is possible to provide many matrices G_1, \dots, G_k and weights $\alpha_1, \dots, \alpha_k$ to Gromov-Wasserstein problems. Finally, in the case of cortical functional alignment, one could be interested solving a combination of all previously mentioned constraints.

Part II

Contributions

Chapter 3

FUGW: an optimal transport method to align cortical surfaces using anatomical and functional data

3.1 Introduction

The availability of millimeter or sub-millimeter anatomical or functional brain images has opened new horizons to neuroscience, namely that of mapping cognition in the human brain and detecting markers of diseases. Yet this endeavour has stumbled on the roadblock of inter-individual variability: while the overall organization of the human brain is largely invariant, two different brains (even from monozygotic twins [[Pizzagalli et al., 2020](#)]) may differ at the scale of centimeters in shape, folding pattern, and functional responses. The problem is further complicated by the fact that functional images are noisy, due to imaging limitations and behavioural differences across individuals that cannot be easily overcome. The status quo of the field is thus to rely on anatomy-based inter-individual alignment that approximately matches the outline of the brain [[Avants et al., 2008](#)] as well as its large-scale cortical folding patterns [[Dale et al., 1999](#); [Fischl, 2012](#)]. Existing algorithms thus coarsely match anatomical features with diffeomorphic transformations, by warping individual data to a simplified template brain. Such methods lose much of the original individual detail and blur the functional information that can be measured in brain regions (see [Figure 3.1](#)).

In order to improve upon the current situation, a number of challenges have to be addressed: (i) Commonly used cortical alignment methods are generally blind to functional information, and there are no functional equivalent to

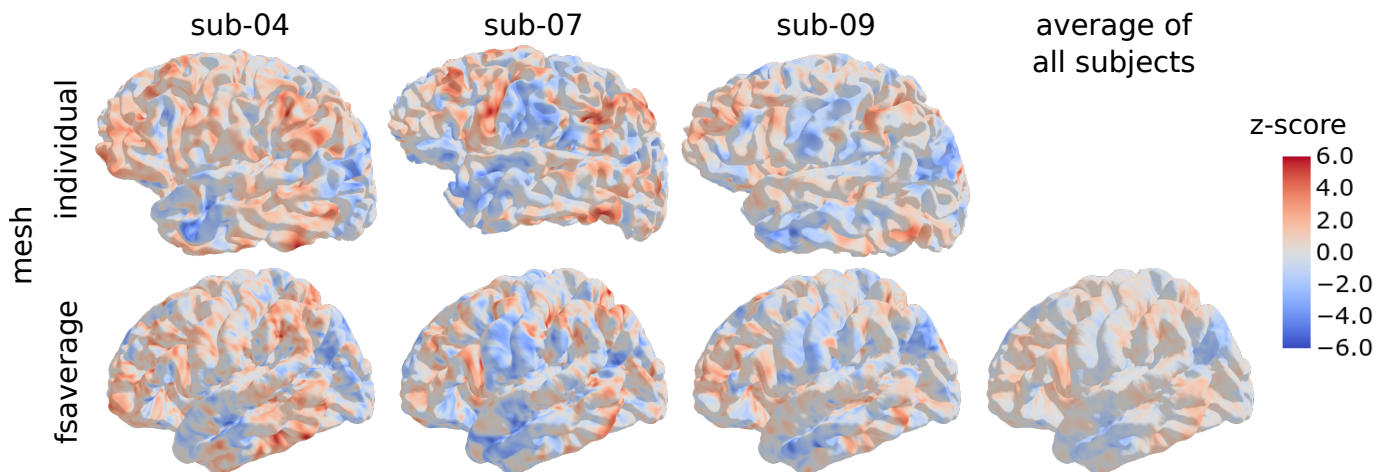


Figure 3.1: High variability in human anatomies and functional MRI responses across subjects In this experiment contrasting areas of the brain which respond to mathematical tasks against other that don't, we observe great variability in locations and strength of brain activations across subjects (row 1). The classical approach consists in warping this data to a common surface template (row 2), where they can be averaged, often resulting in loss of individual details and detection power. These images were generated using Nilearn software [Abraham et al., 2014].

anatomical templates. This is unfortunate, since functional information is arguably the most accessible marker to identify cortical regions and their boundaries [Glasser et al., 2016]. (ii) When comparing two brains – coming from individuals or from a template – it is unclear what regularity should be imposed on the matching [Van Essen et al., 2012]. While it is traditional in medical imaging to impose diffeomorphicity [Avants et al., 2008; Robinson et al., 2014], such a constraint does not match the frequent observation that brain regions vary across individuals in their fine-grained functional organization [Glasser et al., 2016; Schneider et al., 2019]. (iii) Beyond the problem of aligning human brains, it is an even greater challenge to systematically compare functional brain organization in two different species, such as humans and macaques [Eichert et al., 2020; Mars et al., 2018; Neubert et al., 2014; Xu et al., 2020]. Such interspecies comparisons introduce a more extreme form of variability in the correspondence model.

Related work Several attempts have been made to constrain the brain alignment process by using functional information. The first one consists in introducing functional maps into the diffeomorphic framework and search for a smooth transformation that matches functional information [Robinson et al., 2014; Sabuncu et al., 2010; Yeo et al., 2010], the most popular framework being arguably Multimodal Surface Matching (MSM) [Glasser et al., 2016; Robinson et al., 2014].

A second family of less constrained functional alignment approaches have

been proposed, based on heuristics, by matching information in small, possibly overlapping, cortical patches [Bazeille et al., 2021; Haxby et al., 2011; Tavor et al., 2016]. This popular framework has been called *hyperalignment* [Guntupalli et al., 2016; Haxby et al., 2011], or *shared response models* [P-H. Chen et al., 2015]. Yet these approaches lack a principled framework and cannot be considered to solve the matching problem at scale. Neither do they allow to estimate a group-level template properly [Al-Wasity et al., 2020].

An alternative functional alignment framework has followed another path [Gramfort et al., 2015], considering functional signal as a three-dimensional distribution, and minimizing the transport cost. However, this framework imposes unnatural constraints of non-negativity of the signal and only works for one-dimensional contrasts, so that it cannot be used to learn multi-dimensional anatomo-functional structures. An important limitation of the latter two families of methods is that they operate on a fixed spatial context (mesh or voxel grid), and thus cannot be used on heterogeneous meshes such as between two individual human anatomies or, worse, between a monkey brain and a human brain.

Contributions Following Bazeille et al., 2019, we use the Wasserstein distance between source and target functional signals – consisting of contrast maps acquired with fMRI – to compute brain alignments. We contribute two notable extensions of this framework: (i) a Gromov-Wasserstein (GW) term to preserve global anatomical structure – this term introduces an anatomical penalization against improbably distant anatomical matches, yet without imposing diffeomorphic regularity – as well as (ii) an unbalanced correspondence that allows mappings from one brain to another to be incomplete, for instance because some functional areas are larger in some individuals than in others, or may simply be absent. We show that this approach successfully addresses the challenging case of different cortical meshes, and that derived brain activity templates are sharper than those obtained with standard anatomical alignment approaches.

3.2 Methods

Optimal Transport yields a natural framework to address the alignment problem, as it seeks to derive a plan – a *coupling* – that can be seen as a soft assignment matrix between cortical areas of a source and target individual. As discussed previously, there is a need for a functional alignment method that respects the rich geometric structure of the anatomical features, hence the Wasserstein distance alone is not sufficient. By construction, the GW distance [Memoli, 2007; Memoli, 2011] can help preserve the global geometry underlying the signal. The more recent fused GW distance [Vayer et al., 2020] goes one step further by making it possible to integrate functional data simultaneously

with anatomical information.

3.2.1 Fused Unbalanced Gromov-Wasserstein

We leverage Vayer et al., 2020 and Séjourné et al., 2021 to present a new objective function which interpolates between a loss preserving the global geometry of the underlying mesh structure and a loss aligning source and target features, while simultaneously allowing not to transport some parts of the source and target distributions. We provide an open-source solver that minimizes this loss¹.

Formulation We denote $F^s \in \mathbb{R}^{n,c}$ the matrix of features per vertex for the source subject. In the proposed application, they correspond to c functional activation maps, sampled on a mesh with n vertices representing the source subject's cortical surface. Let $D^s \in \mathbb{R}_+^{n,n}$ be the matrix of pairwise geodesic distances² between vertices of the source mesh. Moreover, we assign the distribution $w^s \in \mathbb{R}_+^n$ on the source vertices. Comparably, we define $F^t \in \mathbb{R}^{p,c}$, $D^t \in \mathbb{R}_+^{p,p}$ and $w^t \in \mathbb{R}_+^p$ for the target subject, whose individual anatomy is represented by a mesh comprising p vertices. Eventually, w^s and w^t set the transportable mass per vertex, which, without prior knowledge, we choose to be uniform for the source and target vertices respectively: $w^s \triangleq (\frac{1}{n}, \dots, \frac{1}{n})$, $w^t \triangleq (\frac{1}{p}, \dots, \frac{1}{p})$.

Given a tuple of hyper-parameters $\theta \triangleq (\rho, \alpha, \varepsilon)$, where $\rho, \varepsilon \in \mathbb{R}_+$ and $\alpha \in [0, 1]$, for any coupling $P \in \mathbb{R}^{n,p}$, we define the fused unbalanced Gromov-Wasserstein loss as

$$\begin{aligned}
 L_\theta(P) \triangleq & (1 - \alpha) \sum_{\substack{0 \leq i < n \\ 0 \leq j < p}} \|F_i^s - F_j^t\|_2^2 P_{i,j} + \alpha \sum_{\substack{0 \leq i, k < n \\ 0 \leq j, l < p}} |D_{i,k}^s - D_{j,l}^t|^2 P_{i,j} P_{k,l} \\
 & + \rho \left(\text{KL}(P_{\#1} \otimes P_{\#1} \mid w^s \otimes w^s) + \text{KL}(P_{\#2} \otimes P_{\#2} \mid w^t \otimes w^t) \right) + \varepsilon E(P)
 \end{aligned} \tag{3.1}$$

Wasserstein loss $L_W(P)$ Gromov-Wasserstein loss $L_{GW}(P)$
Marginal constraints $L_U(P)$ Entropic regularisation

where $L_W(P)$ matches vertices with similar features, $L_{GW}(P)$ penalizes changes in geometry and $L_U(P)$ fosters matching all parts of the source and target distributions. Throughout this paper, we refer to relaxing the hard marginal constraints of the underlying OT problem into soft ones as *unbalancing*. Here, $P_{\#1} \triangleq (\sum_j P_{i,j})_{0 \leq i < n}$ denotes the first marginal distribution of P , and $P_{\#2} \triangleq (\sum_i P_{i,j})_{0 \leq j < p}$ the second marginal distribution of P . The notation \otimes represents the Kronecker product between two vectors or two tensors. $\text{KL}(\cdot|\cdot)$ de-

¹<https://github.com/alexisthual/fugw> provides a PyTorch [Paszke et al., 2019] solver with a scikit-learn [Pedregosa et al., 2011] compatible API

²We compute geodesic distances using <https://github.com/the-virtual-brain/tvb-gdist>

notes the Kullback Leibler divergence, which is a typical choice to measure the discrepancy between two measures in the context of unbalanced optimal transport [Liero et al., 2018]. The last term $E(P) \triangleq \text{KL}(P \otimes P \mid (w^s \otimes w^t) \otimes (w^s \otimes w^t))$ is mainly introduced for computational purposes, as it helps accelerate the approximation scheme of the optimisation problem. Typically, it is used in combination with a small value of ε , so that the impact of other terms is not diluted. On the other hand, the parameters α and ρ offer control over two other aspects of the problem: while α realizes a trade-off between the impact of different features and different geometries in the resulting alignment, ρ controls the amount of mass transported by penalizing configurations such that the marginal distributions of the transportation plan P are far from the prior weights w^s and w^t . This potentially helps adapt the size of areas where either the signal or the geometry differs too much between source and target.

Eventually, we define $\mathcal{X}^s \triangleq (F^s, D^s, w^s)$ and $\mathcal{X}^t \triangleq (F^t, D^t, w^t)$, and seek to derive an optimal coupling $P \in \mathbb{R}^{n,p}$ minimizing

$$\text{FUGW}(\mathcal{X}^s, \mathcal{X}^t) \triangleq \inf_{P \geq 0} L_\theta(P) \quad (3.2)$$

This can be seen as a natural combination of the fused GW [Vayer et al., 2020] and the unbalanced GW [Séjourné et al., 2021] distances. To the best of our knowledge, it has never been considered in the literature.

Toy example illustrating the unbalancing property As exemplified in Figure 3.1, brain responses elicited by the same stimulus vary greatly between individuals. Figure 3.2 illustrates a similar yet simplified version of this problem, where the goal is to align two different signals supported on the same spherical meshes. In this example, for each of the $n = p = 3200$ vertices, the feature is simply a scalar. On the source mesh, the signal is constituted of two von Mises density functions that differ by their concentration (large and small), while on the target mesh, only the large one is present, but at a different location. We use the optimal coupling matrix P obtained from Eq. 3.2 to transport the source signal on the target mesh. As shown in Figure 3.2.B, the parameter ρ allows to control the mass transferred from source to target. When $\rho = 100$, we approach the solution of the fused GW problem. Consequently, we observe the second mode on the target when transporting the source signal. When the mass control is weaker ($\rho = 1$), the smaller blob is partly removed because it has no counterpart in the target configuration, making the transport ill-posed.

Barycenters Barycenters represent common patterns across samples. Their role is instrumental in identifying a unique target for aligning a given group of individuals. As seen in Fig. 3.1, the vertex-wise group average does not usually provide well-contrasted maps. Inspired by the success of the GW distance when estimating the barycenter of structured objects [Peyré et al., 2016; Vayer

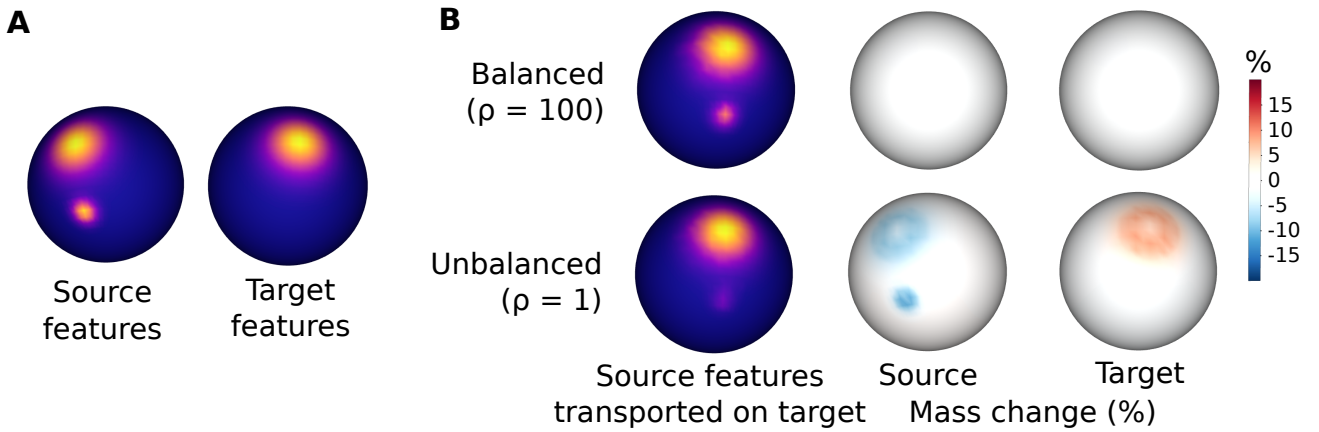


Figure 3.2: Unbalancing helps accounting for idiosyncrasies of the source and target signals When trying to align the source and target signals (Panel A), the classical balanced setup (Panel B, top row) transports all parts of the source signal even if they have no counterpart in the target signal. In the unbalanced setup (Panel B, bottom row), less source-only signal is transported: in particular, less mass is transported from the source’s small blob onto the target (Panel B, middle column).

et al., 2020], we use FUGW to find the barycenter $(F^B, D^B) \in \mathbb{R}^{k,c} \times \mathbb{R}^{k,k}$ of all subjects $s \in \mathcal{S}$, as well as the corresponding couplings $P^{s,B}$ from each subject to the barycenter. More precisely, we solve

$$\mathcal{X}^B = (F^B, D^B, w^B) \in \arg \min_{\mathcal{X}} \sum_{s \in \mathcal{S}} \text{FUGW}(\mathcal{X}^s, \mathcal{X}), \quad (3.3)$$

where we set the weights w_B to be the uniform distribution. By construction, the resulting barycenter benefits from the advantages of FUGW, i.e. equilibrium between geometry-preserving and feature-matching properties, while not forcing hard marginal constraints. The FUGW barycenter is estimated using a Block-Coordinate Descent (BCD) algorithm that consists in alternatively (i) minimizing the OT plans $P^{s,B}$ for each FUGW computation in equation 3.3 with fixed \mathcal{X}^B and (ii) updating the barycenter \mathcal{X}^B through a closed form with fixed $P^{s,B}$. See Algorithm 3.4 for more details.

3.2.2 Approximating solutions to the FUGW problem by minimizing a lower bound

This section provides extensive information as to how we derive solutions to the FUGW problem. It can be skipped at first, but is needed to understand some of the extensions presented in Chapter 4.

Estimating the unbalanced Gromov Wasserstein (UGW) loss is numerically sensitive to initialisation, due to the non-convexity of the problem. Therefore FUGW is *a priori* non-convex as well, and comparably difficult to estimate. Consequently, following Séjourné et al., 2021, we instead compute a lower bound

which we formulate as a bi-convex problem that relies on the joint estimation of two couplings. It involves solving a minimization problem with respect to two independent couplings: using a Block-Coordinate Descent (BCD) scheme, we fix a coupling and minimize with respect to the other. This allows us to always deal with linear problems instead of a quadratic one. Eventually, each BCD iteration consists in alternatively solving two entropic unbalanced OT problems, whose solutions can be approximated using the scaling algorithm [Chizat et al., 2019]. We provide next a detailed derivation of this estimation procedure, using notations introduced in section 3.2.1.

First of all, let us rewrite L_θ , the loss function introduced in Equation 3.1, to eliminate quadratic terms involving P . Let $P, Q \in \mathbb{R}^{n,p}$ be two couplings. We turn each of the terms L_W, L_{GW}, L_U and E of the loss function into a function of P and Q . The updated loss function reads:

$$L_\theta^{\text{bi}}(P, Q) \triangleq (1 - \alpha) L_W(P, Q) + \alpha L_{GW}(P, Q) + \rho L_U(P, Q) + \varepsilon E(P, Q) \quad (3.4)$$

where

$$L_W(P, Q) \triangleq \langle C, \frac{P+Q}{2} \rangle = \frac{1}{2} \left(\sum_{\substack{0 \leq i < n \\ 0 \leq j < p}} C_{i,j} P_{i,j} + \sum_{\substack{0 \leq i < n \\ 0 \leq j < p}} C_{i,j} Q_{i,j} \right)$$

$$L_{GW}(P, Q) \triangleq \langle G, P \otimes Q \rangle = \sum_{\substack{0 \leq i, k < n \\ 0 \leq j, l < p}} G_{i,j,k,l} P_{i,j} Q_{k,l}$$

$$L_U(P, Q) \triangleq \text{KL}(P_{\#1} \otimes Q_{\#1} \mid w^s \otimes w^s) + \text{KL}(P_{\#2} \otimes Q_{\#2} \mid w^t \otimes w^t)$$

$$E(P, Q) \triangleq \text{KL}(P \otimes Q \mid (w^s \otimes w^t) \otimes (w^s \otimes w^t))$$

and

$$C \triangleq \left(\|F_i^s - F_j^t\|_2^2 \right)_{i,j} \in \mathbb{R}_+^{n,p} \text{ is the feature cost matrix,}$$

$$G \triangleq \left(|D_{i,k}^s - D_{j,l}^t|^2 \right)_{i,j,k,l} \in \mathbb{R}_+^{n,p,n,p} \text{ is the geometry cost tensor.}$$

It is important to note that, when $P = Q$, the objective functions of Equation 3.4 and Equation 3.1 are equal: $L_\theta^{\text{bi}}(P, P) = L_\theta(P)$. Therefore, solving for the new loss under the constraint $P = Q$ is the same problem as solving for the original FUGW loss. We relax this equality constraint between the two couplings and only require that the mass of P and Q be equal, where the mass of a vector is defined as the sum of its coefficients: $m(P) \triangleq \sum_{i,j} P_{i,j}$. Consequently, we define the following lower bound for the FUGW problem:

$$\text{FUGW}(\mathcal{X}^s, \mathcal{X}^t) = \inf_{\substack{P, Q \geq 0 \\ P=Q}} L_\theta^{\text{bi}}(P, Q) \geq \inf_{\substack{P, Q \geq 0 \\ m(P)=m(Q)}} L_\theta^{\text{bi}}(P, Q) \triangleq \text{LB-FUGW}(\mathcal{X}^s, \mathcal{X}^t) \quad (3.5)$$

The tightness of this lower bound is studied in [Séjourné et al., 2021](#) for the unbalanced Gromov-Wasserstein problem but is out of the scope of our work here.

Solver We provide a Python GPU-based solver for LB-FUGW, using an approach similar to that of [Séjourné et al., 2021](#), which we recall in Algorithm 3.1. More precisely, we alternatively optimize one coupling while keeping the other fixed. It is possible because [Séjourné et al., 2021](#) showed that when Q is fixed, minimizing a regularised unbalanced Gromov-Wasserstein loss with respect to P reads as a regularised unbalanced Wasserstein problem, where the associated cost matrix K is a function of Q and the other terms of the loss function – this derives from calculations made in the proofs of Propositions 9 and 10 from the appendix of [Séjourné et al., 2021](#). Fortunately, plenty of solvers exist for this optimisation problem. Respectively, this result holds when P is fixed and one optimises for Q .

We extend this result by showing that when Q is fixed, the minimization of L_θ with respect to P also reads as a regularised unbalanced Wasserstein problem, where the associated cost matrix K is a new function of Q and the other terms of the loss function. In the interest of brevity, we do not report these calculations here, but the interested reader will find similar ones in Section 4.7.

Besides, in Algorithm 3.2, we detail the computation of the local cost K used in the previous algorithm. It explicitly shows actual computations we make on CPUs and GPUs to compute the local cost matrix, which exploits the underlying structure of G to maximise computational efficiency. Note that we do not store the full cost tensor G , which lowers the memory footprint of this approach. Finally, this results in two entropic unbalanced OT problems in each iteration, which can be solved using the scaling algorithm from [Chizat et al., 2019](#) detailed in Algorithm 3.3.

Algorithm 3.1 Approximation scheme for LB-FUGW

Input: $\mathcal{X}^s, \mathcal{X}^t, \alpha, \rho, \varepsilon$

Output: Pair of optimal couplings (P, Q)

- 1: initialise: $P = Q = w^s \otimes w^t / \sqrt{m(w^s) m(w^t)}$
 - 2: **while** (P, Q) has not converged **do**
 - 3: $K_P \leftarrow \text{Cost}(P, G, C, w^s, w^t, \alpha, \rho, \varepsilon)$ ▷ fixed P
 - 4: $Q \leftarrow \text{Scaling}(K_P, w^s, w^t, \rho m(P), \varepsilon m(P))$
 - 5: $Q \leftarrow \sqrt{\frac{m(P)}{m(Q)}} Q$
 - 6: $K_Q \leftarrow \text{Cost}(Q, G, C, w^s, w^t, \alpha, \rho, \varepsilon)$ ▷ fixed Q
 - 7: $P \leftarrow \text{Scaling}(K_Q, w^s, w^t, \rho m(Q), \varepsilon m(Q))$
 - 8: $P \leftarrow \sqrt{\frac{m(Q)}{m(P)}} P$
 - 9: **end while**
-

Algorithm 3.2 Compute local bi-convex cost matrix from LB-FUGW

Input: $P, G, C, w^s, w^t, \alpha, \rho, \varepsilon$ **Output:** Local cost K

1:

$$\begin{aligned}
 K &\leftarrow \frac{1-\alpha}{2} C \\
 &+ \alpha \left[\left((D^s \odot D^s) P_{\#1} \oplus (D^t \odot D^t) P_{\#2} \right) - 2D^s P (D^t)^T \right] \\
 &+ \rho \left[\left\langle \log \frac{P_{\#1}}{w^s}, P_{\#1} \right\rangle + \left\langle \log \frac{P_{\#2}}{w^t}, P_{\#2} \right\rangle \right] \mathbb{1}_n \otimes \mathbb{1}_p \\
 &+ \varepsilon \left[\left\langle \log \frac{P}{w^s \otimes w^t}, P \right\rangle \right] \mathbb{1}_n \otimes \mathbb{1}_p
 \end{aligned}$$

Algorithm 3.3 Scaling algorithm from [Chizat et al., 2019](#)

Input: $K, w^s, w^t, \rho, \varepsilon$ **Output:** Optimal coupling P 1: initialise dual vectors: $u = \mathbb{0}_n, v = \mathbb{0}_p$ 2: **while** (u, v) has not converged **do**3: $u \leftarrow -\frac{\rho}{\rho+\varepsilon} \log \left(\sum_j \exp \left((v_j + \log w_j^t) \mathbb{1}_n - \frac{K_{\cdot,j}}{\varepsilon} \right) \right)$ 4: $v \leftarrow -\frac{\rho}{\rho+\varepsilon} \log \left(\sum_i \exp \left((u_i + \log w_i^s) \mathbb{1}_p - \frac{K_{i,\cdot}}{\varepsilon} \right) \right)$ 5: **end while**6: $P \leftarrow (w^s \otimes w^t) \odot \exp \left(u \oplus v - \frac{K}{\varepsilon} \right)$

Here, the \odot denotes the Hadamard, i.e., the element-wise product of two matrices. Similarly, the exponential, division, and logarithm operations are computed element-wise. The scalar product is denoted by $\langle \cdot, \cdot \rangle$.

In practice, we observe that the two couplings of LB-FUGW are numerically equal. Therefore, we only use P – and not Q .

3.2.3 Detailed description of FUGW barycenter estimation

FUGW-barycenter algorithm, described in Algorithm 3.4, alternates between computing mappings from subjects to the barycenter, and updating the barycenter. This corresponds to a block coordinate descent on the barycenter estimation. The first step simply uses the previously introduced solver. The second one takes advantage of the fact that the objective function introduced in Equation 3.5 is differentiable in F^B and D^B , and the two couplings of LB-FUGW are numerically equal. This yields a closed form for F^B and D^B , as a function of $P^{s,B}$ and \mathcal{X}^s . We note that, during the barycenter estimation, the weight w^B is always fixed as uniform distribution. Finally, it might be costly to iterate over all subjects at each step. Therefore, we suggest randomly drawing a subset of subjects at each iteration, which we denote by \tilde{S} .

Algorithm 3.4 LB-FUGW barycenter

Input: $(\mathcal{X}^s)_{s \in \mathcal{S}}, \alpha, \rho, \varepsilon$

Output: Individual couplings $(P^{s,B})_{s \in \mathcal{S}}$, barycenter \mathcal{X}^B

- 1: initialise: $F^B = \mathbb{1}_k; D^B = \mathbb{0}_k$
- 2: **while** $\mathcal{X}^B = (F^B, D^B, w^B)$ has not converged **do**
- 3: Draw \tilde{S} subset of S
- 4: **for** $s \in \tilde{S}$ **do** ▷ fixed \mathcal{X}^B
- 5: Align: $P^{s,B} \leftarrow \text{LB-FUGW}(\mathcal{X}^s, \mathcal{X}^B, \rho, \alpha, \varepsilon)$
- 6: **end for**
- 7: Update F^B and D^B : ▷ fixed $P^{s,B}$

$$F^B \leftarrow \frac{1}{|\tilde{S}|} \sum_{s \in \tilde{S}} \text{diag} \left(\frac{1}{P_{\#2}^{s,B}} \right) (P^{s,B})^\top F^s$$

$$D^B \leftarrow \frac{1}{|\tilde{S}|} \sum_{s \in \tilde{S}} \frac{(P^{s,B})^\top D^s P^{s,B}}{P_{\#2}^{s,B} (P_{\#2}^{s,B})^\top}$$

- 8: **end while**
-

3.3 Numerical experiments

We design three experiments to assess the performance of FUGW. In Experiments 1 and 2, we are interested in assessing if aligning pairs of individuals with FUGW increases correlation between subjects compared to a baseline correlation. We also compare the ensuing gains with those obtained when using the competing method MSM [Robinson et al., 2014, 2018] to align subjects. In Experiment 3, we derive a barycenter of individuals and assess its ability to capture fine-grained details compared to classical methods.

Dataset We leverage data from the Individual Brain Charting dataset [Pinho et al., 2018] in all three experiments. It is a longitudinal study on 12 human subjects, comprising 400 fMRI maps per subject collected on a wide variety of stimuli (motor, visual, auditory, theory of mind, language, mathematics, emotions, and more), movie-watching data, T1-weighted maps, as well as other features such as retinotopy which we don't use in this work. We leverage these 400 fMRI maps. The training, validation and test sets respectively comprise 326, 43 and 30 contrast maps acquired for each individual of the dataset. Tasks and MRI sessions differ between each of the sets. More details, including preprocessing, are provided in Supplementary Materials.

Baseline alignment correlation For each pair of individuals (s, t) under study, and for each fMRI contrast c in the test set, we compute the Pearson correlation $\text{corr}(F_{\cdot,c}^s, F_{\cdot,c}^t)$ after these maps have been projected onto a common surface anatomy (in this case, *fsaverage5* mesh). Throughout this work, such computations are made for each hemisphere separately.

Experiment 1 - Aligning pairs of humans with the same anatomy For each pair (s, t) under study, we derive an alignment $\mathbf{P}^{s,t} \in \mathbb{R}^{n \times p}$ using FUGW on a set of training features. In this experiment, source and target data lie on the same anatomical mesh (*fsaverage5*), and $n = p = 10\,240$ for each hemisphere. Since each hemisphere’s mesh is connected, we align one hemisphere at a time.

Computed couplings are used to align contrast maps of a the validation set from the source subject onto the target subject. Indeed, one can define

$$\phi_{s \rightarrow t}: \mathbf{X} \in \mathbb{R}^{n \times q} \mapsto ((\mathbf{P}^{s,t})^T \mathbf{X}) \oslash \mathbf{P}_{\#2}^{s,t} \in \mathbb{R}^{p \times q} \quad (3.6)$$

where \oslash represents the element-wise division. $\phi_{s \rightarrow t}$ transports any matrix of features from the source mesh to the target mesh. We measure the Pearson correlation $\text{corr}(\phi_{s \rightarrow t}(\mathbf{F}^s), \mathbf{F}^t)$ between each aligned source and target maps.

We run a similar experiment for MSM and compute the correlation gain induced on a test set by FUGW and MSM, respectively. We selected the hyper-parameters maximizing correlation gain for both models on a validation set. In the case of FUGW, in addition to gains in correlation, hyper-parameter selection was influenced by three other metrics that help us assess the relevance of computed couplings:

Transported mass For each vertex i of the source subject, we compute

$$\sum_{0 \leq j < p} \mathbf{P}_{i,j}^{s,t}$$

the coefficient of the marginal associated with vertex i

Vertex displacement Taking advantage of the fact that the source and target anatomies are the same, we define $\mathbf{D} = \mathbf{D}^s = \mathbf{D}^t$ and compute for each vertex i of the source subject the quantity $\sum_j \mathbf{P}_{i,j}^{s,t} \cdot \mathbf{D}_{i,j} / \sum_j \mathbf{P}_{i,j}^{s,t}$, which measures the average geodesic distance on the cortical sheet between vertex i and the vertices of the target it has been matched with

Vertex spread Large values of ε increase the entropy of derived couplings. To quantify this effect, and because we don’t want the matching to be too blurry, we assess how much a vertex was *spread*. Considering $\tilde{P}_i = \mathbf{P}_i^{s,t} / \sum_j \mathbf{P}_{i,j}^{s,t} \in \mathbb{R}^p$ as a probability measure on target vertices, we estimate the anatomical variance of this measure by sampling q pairs (j_q, k_q) of \tilde{P}_i and computing their average geodesic distance $\frac{1}{q} \sum_{j_q, k_q} \mathbf{D}_{j_q, k_q}$

Experiment 2 - Aligning pairs of humans with individual anatomies We perform a second alignment experiment, this time using individual meshes instead of an anatomical template.

However, individual meshes are significantly larger than the common anatomical template used in Experiment 1 ($n \approx m \approx 160\text{k}$ vs. 10k previously), resulting in couplings too large to fit on GPUs – for reference, a coupling of size $10\text{k} \times 10\text{k}$ already weights 400Mo on disk. We thus reduce the size of the source and target data by clustering them into 10k small connected clusters using Ward’s

algorithm [B. Thirion et al., 2014]. More details are given in supplementary section A.4.

Experiment 3 - Comparing FUGW barycenters with usual group analysis

Since it is very difficult to estimate the barycentric mesh, we force it to be equal to the *fsaverage5* template. Empirically, this we force the distance matrix D^B to be equal to that of *fsaverage5*, and only estimate the functional barycenter F^B . We initialise it with the mean of $(F^s)_{s \in S}$ and derive F^B and $(P^{s,B})_{s \in S}$ from problem 3.3.

Then, for a given stimulus c , we compute its projection onto the barycenter for each subject. We use these projections to compute two maps of interest: (i) $M_{B,c}$ the mean of projected contrast maps across subjects and (ii) $T_{B,c}$ the t-statistic (for each vertex) of projected maps. We compare these two maps with their unaligned counterparts $M_{0,c}$ and $T_{0,c}$ respectively.

$$M_{B,c} \triangleq \frac{1}{|S|} \sum_{s \in S} \phi_{s \rightarrow t}(F^{s,c}) \quad T_{B,c} \triangleq \text{t-statistic}\left(\left(\phi_{s \rightarrow t}(F^{s,c})\right)_{s \in S}\right)$$

$$M_{0,c} \triangleq \frac{1}{|S|} \sum_{s \in S} F^{s,c} \quad T_{0,c} \triangleq \text{t-statistic}\left(\left(F^{s,c}\right)_{s \in S}\right)$$

The first map helps us to qualitatively evaluate the precision of FUGW alignments and barycenter. The second one is classically used to infer the existence of areas of the brain that respond to specific stimuli. We assess whether FUGW helps find the same clusters of vertices. Eventually, we quantify the number of vertices significantly activated or deactivated with and without alignment respectively.

3.4 Results

3.4.1 Experiment 1 - Template anatomy

Aligning subjects on a fixed mesh We set $\alpha = 0.5$, $\rho = 1$ and $\varepsilon = 10^{-3}$. Pearson correlation between source and target contrast maps is systematically and significantly increased when aligned using FUGW, as illustrated in Figure 3.3 where correlation grows by almost 40% from 0.258 to 0.356.

We also varied training sets by selecting subsets of training contrasts and find that similar performance on the test set can be achieved regardless of the training data (see Supplementary section 9.3 and in particular Supplementary Table 9.1).

Hyper-parameters selection Hyper-parameters used to obtain these results were chosen after running a grid search on α , ε and ρ and evaluating it on the validation dataset. Computation took about 100 hours using 4 Tesla V100-DGXS-

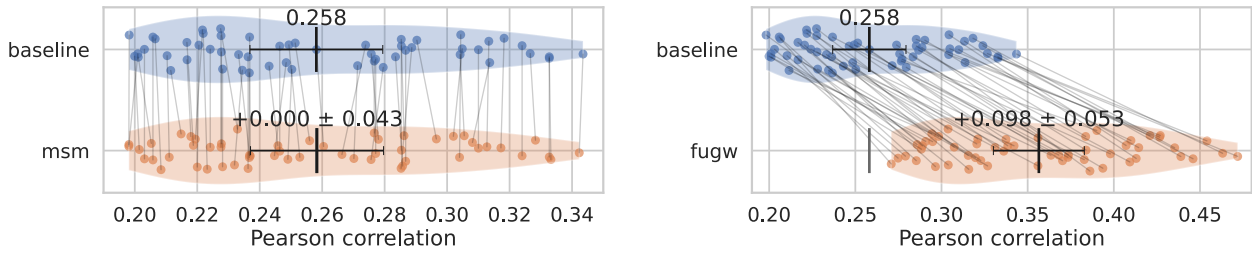


Figure 3.3: Comparison of gains in correlation after inter-subject alignment

For each pair of source and target subjects of the dataset, we compute the average Pearson correlation between 30 test contrasts, leading to the (baseline) correspondence score, and compare it with that of the same contrast maps aligned with either MSM (left) or FUGW (right). Correlation gains are much better for FUGW.

32GB GPUs. More precisely, it takes about 4 minutes to compute one coupling between a source and target 10k-vertex hemisphere on a single GPU, when the solver was set to run 10 BCD and 400 Sinkhorn iterations. In comparison, MSM takes about the same time on Intel(R) Xeon(R) CPU E5-2698 v4 @ 2.20GHz CPUs. Results are reported in Figure 3.4 and provide multiple insights concerning FUGW.

Firstly, without anatomical constraint ($\alpha = 0$), source vertices can be matched with target vertices that are arbitrarily far on the cortical sheet. Even though this can significantly increase correlation, it also results in very high vertex displacement values (up to $100mm$). Such couplings are not anatomically plausible. Secondly, without functional information ($\alpha = 1$), couplings recover a nearly flawless matching between source and target meshes, so that, when $\varepsilon = 10^{-5}$ (ie when we force couplings to find single-vertex-to-single-vertex matches), vertex displacement and spread are close to 0 and correlation is unchanged. Fusing both constraints ($0 < \alpha < 1$) yields the largest gains in correlation while allowing to compute anatomically plausible reorganizations the cortical sheet between subjects.

The impact of ρ (controlling marginal penalizations) on correlation seems modest, with a slight tendency of increased correlation in unbalanced problems (low ρ).

Finally, it is worth noting that a relatively wide range of α and ρ yield comparable gains. The fact that FUGW performance is weakly sensitive to hyper-parameters makes it a good off-the-shelf tool for neuroscientists who wish to derive inter-individual alignments. However, ε is of dramatic importance in computed results and should be chosen carefully. Vertex spread is a useful metric to choose sensible values of ε ; for human data one might consider that it should not exceed $20mm$.

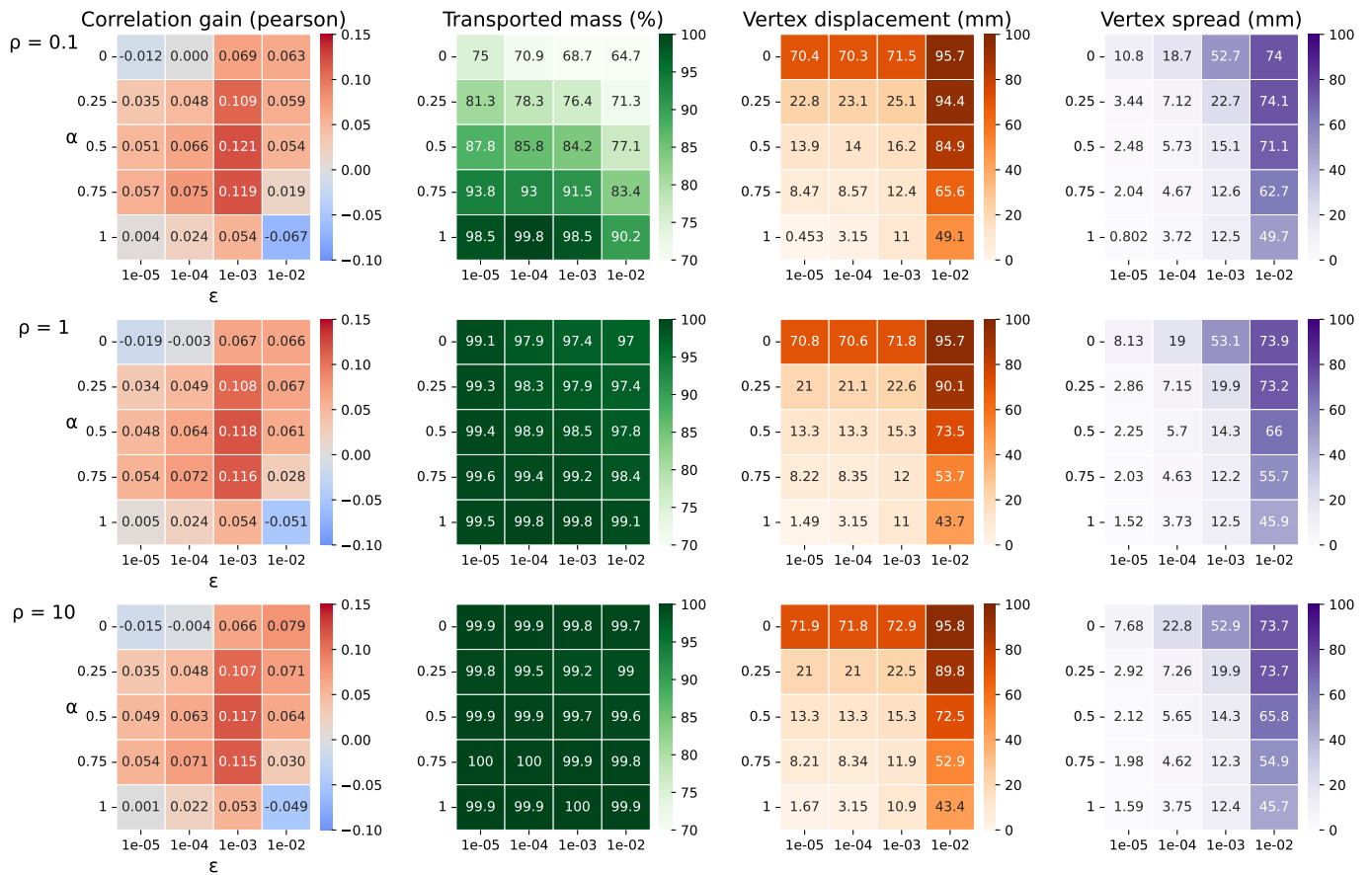


Figure 3.4: Exploring hyper-parameter space to find relevant couplings Given a transport plan aligning a source and target subject, we evaluate how much this coupling (left) improves correlation between unseen contrast maps of the two subjects, (center left) actually transports data, (center right) moves vertices far from their original location on the cortical surface and (right) spreads vertices on the cortical sheet. We seek plans that maximize correlation gain, while keeping spread and displacement low enough.

Mass redistribution in unbalanced couplings Unbalanced couplings provide additional information about how functional areas might differ in size between pairs of individuals. This is illustrated in Figure 3.5, where we observe variation in size of the auditory area between a given pair of individuals. This feature is indeed captured by the difference of mass between subjects (although the displayed contrast was not part of the training set).

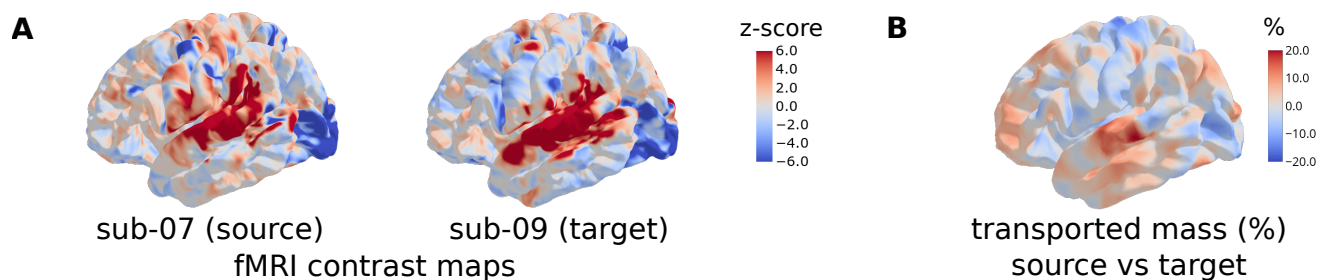


Figure 3.5: Transported mass indicates areas which have to be resized between subjects (Panel A) We show a contrast map from the test set which displays areas showing stronger activation during auditory tasks versus equivalent visual tasks. It shows much more anterior activations on the target subject compared to the source subject. This is consistent with the observation that more mass is present in anterior auditory areas of the source subject than in the target subject (Panel B).

3.4.2 Experiment 2 - Individual anatomies

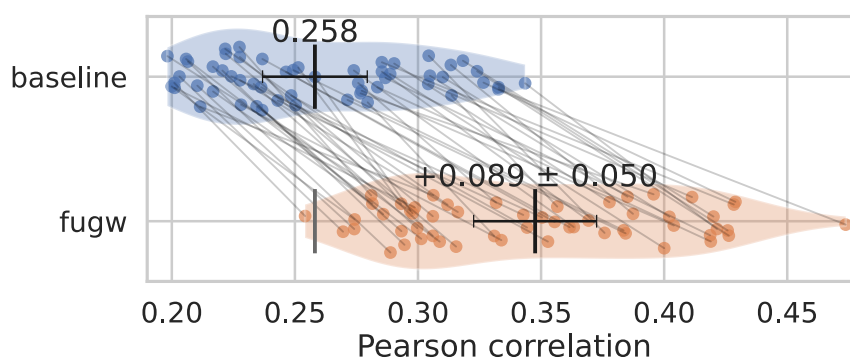


Figure 3.6: Correlation between pairs of subjects is significantly better after alignment on individual anatomies than after projecting subjects onto a common anatomical template

As shown in Figure 3.6, we obtain correlation gains which are comparable to that of Experiment 1 (about 35% gain) while working on individual meshes. This tends to show that FUGW can compute meaningful alignments between pairs of individuals without the use of an anatomical template, which helps bridge most conceptual impediments listed in Section 3.1.

Moreover, this opens the way for computation of simple statistics in cohorts

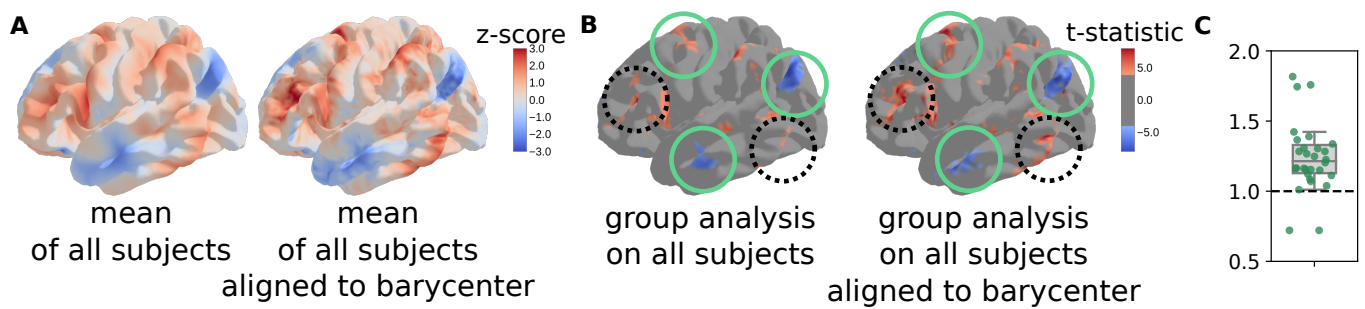


Figure 3.7: FUGW barycenter yields much finer-grained maps than group averages We study the same statistical map as in Figure 3.1, which contrasts areas of the brain involved in mathematical reasoning. **A.** These complex maps projected onto the barycenter and averaged show more specific activation patterns than simple group averages – average functional response in each vertex of the anatomical template – especially in cortical areas exhibiting more variability, such as the prefrontal cortex. **B.** Deriving a t-test on aligned maps captures the same clusters as the classical approach (plain green circles), but also new clusters in areas where inter-subject variability is high (dotted black circles). Peak t-statistics are also higher with FUGW. **C.** Ratio of number of activated vertices ($|t\text{-statistic}| \geq 4$) with versus without alignment for each map of the test set. Our method finds significantly more of such vertices ($p\text{-value} = 3 \cdot 10^{-4}$).

of individuals in the absence of a template. Indeed, one can pick an individual of the cohort and use it as a reference subject on which to transport all other individuals. We give an example in Figure 9.4, showing that FUGW correctly preserved idiosyncrasies of each subject while transporting their functional signal in an anatomically sound way.

3.4.3 Experiment 3 - Barycenter

In the absence of a proper metric to quantify the correctness of a barycenter, we first qualitatively compare the functional templates obtained with and without alignment. In Figure 3.7.A, we do so using brain maps taken from the test set. We can see that the barycenter obtained with FUGW yields sharper contrasts and more fine-grained details than the barycenter obtained by per-vertex averaging. We also display in Figure 3.7.B the result of a one-sample test for the same contrast, which can readily be used for inference. The one-sample test map obtained after alignment to the FUGW template exhibits the same supra-threshold clusters as the original approach, but also some additional spots which were likely lost due to inter-subject variability in the *fsaverage5* space. This approach is thus very useful to increase power in group inference. We quantify this result by counting the number of supra-threshold vertices with and without alignment for each contrast map of the test set. Our alignment method recovers a much higher number of significantly activated vertices in group analysis, as shown in Figure 3.7.C.

3.5 Discussion

FUGW can derive meaningful couplings between pairs of subjects without the need of a pre-existing anatomical template. It is well-suited to computing barycenters of individuals, even for small cohorts.

In addition, we have shown clear evidence that FUGW yields gains that cannot be achieved by traditional diffeomorphic registration methods. These methods impose very strong constraints to the displacement field, that may prevent reaching optimal configurations. More deeply, this finding suggests that brain comparison ultimately requires lifting hard regularity constraints on the alignment models, and that two human brains differ by more than a simple continuous surface deformation. However, current results have not shown a strong correlation gain of unbalanced OT compared to balanced OT, likely because the cohort under study is too small. Leveraging datasets such as HCP [Van Essen, Smith, Barch, Behrens, Yacoub, Ugurbil et al., 2013] with a larger number of subjects will help lower the standard error on correlation gain estimates. In this work, we decided to rely on a predefined anatomical template (*fsaverage5*) to derive functional barycenters. It would be interesting to investigate whether more representative anatomical templates can be learned during the process. This would in particular help to customize templates to different populations or species. Additionally, using an entropic solver introduces a new hyper-parameter ε that has a strong effect, but is hard to interpret. Future work may replace the scaling algorithm [Chizat et al., 2019] used here by the majorization-minimization one [Chapel et al., 2021], which does not require entropic smoothing. This solution can yield sparse couplings while being orders of magnitude faster, which will prove useful when computing barycenters on large cohorts.

Finally, we plan to make use of FUGW to derive alignments between human and non-human primates without anatomical priors. Indeed, the understanding of given brain mechanisms will benefit from more detailed invasive measurements made on other species *only if* brains can be matched across species; moreover, this raises the question of features that make the human brain unique, by identifying patterns that have no counterpart in other species. By maximizing the functional alignment between areas, but also allowing for some regions to be massively shrunk or downright absent in one species relative to the other, the present tool could shed an objective light on the important issue of whether and how the language-related areas of the human cortical sheet map onto the architecture of non-human primate brains.

Chapter 4

FUGW extensions

In the previous chapter, we introduced a solver based on Sinkhorn’s algorithm to compute solutions to an optimal transport problem optimising for the FUGW loss. However, many theoretical and computational challenges still remain at this stage. In this Chapter, we present extensions to FUGW which tackle these problems.

First of all, we show that FUGW can be used with training data of limited size, and in particular with naturalistic stimuli. This is important since it is unlikely that a large amount of data will systematically be available for each individual to be aligned. To this end, we train FUGW alignments on the same individuals as in Chapter 3, and use the same test data.

Secondly, we investigate the dependence of FUGW on the initialisation of the solver. We empirically show that not every initialisation strategy is suitable for FUGW, but that the one used in the previous chapter is robust.

Thirdly, on top of a stabilised Sinkhorn solver, we introduce two additional solvers, hoping they would be empirically faster. Computational speed is important since we strive to be able to align a lot of pairs of individuals.

Then, since the size of transport plans grows quadratically with the number of vertices of the source and target brains, the previous method cannot be applied as is to high-resolution data. We show that it is possible to scale FUGW to high-resolution datasets, by adapting the two solvers introduced in the previous section.

Finally, we derive a new solver capable of solving FUGW problems in which the Kullback-Leibler divergence is replaced by the squared L2-norm. This is important since the Kullback-Leibler divergence is not always suitable for the problem at hand: it yields transport plans whose absolute transported mass does not vary much between individuals. Empirically, we find that the squared L2-norm allows for higher absolute differences in transported mass between vertices.

Most of the work presented in this chapter was published in the form of posters at CCN 2022, OHBM 2023 and CCN 2023. All code updates were pushed to the

FUGW GitHub repository¹.

4.1 Aligning human subjects with short acquisition-time fMRI training data

The work presented in this section was published in the form of a short article and poster at CCN 2022. It is a follow-up on the work presented in Chapter 3 in which different types of data are used to train cortical alignments. In particular, we show that a much lower amount of fMRI contrast maps can be used to train FUGW. Besides, we show that naturalistic stimuli, which are potentially easier to acquire than elaborate tasks, can be used as well.

4.1.1 Introduction

Cortical folding, as well as functional patterns of activity observed using fMRI vary a lot between individuals. In order to derive statistics at the population level, neuroscientists commonly rely on coarse anatomical information so as to co-register individuals [Avants et al., 2008; Dale et al., 1999; Fischl, 2012].

Several methods have been developed to integrate functional data when deriving inter-subject alignments. Most prominently, Multimodal Surface Matching (MSM) [Glasser et al., 2016; Robinson et al., 2014] computes a diffeomorphic transformation between cortical surfaces of pairs of individuals to match vertices with similar functional activity. However, diffeomorphisms computed with MSM can't capture the fact that cortical areas vary in size and organization between subjects. Other methods like *hyperalignment* [Haxby et al., 2011] or *shared response models* [P.-H. Chen et al., 2015] rely solely on functional data to compute alignments. Nevertheless, the two latter methods don't prevent matching vertices which are arbitrarily distant on the cortex and may therefore lead to anatomically implausible pairings.

Following the work of Bazeille et al., 2019, we leveraged optimal transport to derive a new alignment method denoted as FUGW and detailed in Thual et al., 2022. It derives inter-subject alignments that take advantage of functional data while trying to preserve the global anatomical geometry of the cortex. We show in Thual et al., 2022 that these alignments greatly increase between-subject correlation of new fMRI data – acquired on tasks and sessions which differ from the training set. However, training datasets used in these experiments are massive (50+ hours of fMRI data resulting in 350+ contrast maps were recorded per subject). We replicate these experiments with training datasets consisting of movie-watching data or fMRI contrasts. They take a maximum of 120 minutes to acquire per subject. We show that these affordable datasets can already

¹<https://github.com/alexisthual/fugw>

help derive anatomically convincing alignments that greatly increase between-subject correlation.

4.1.2 Methods

Using a set of training features, we derive alignments between a source and a target individual. We use them to align new features and finally measure how much between-subject correlation increases after alignment.

4.1.2.1 Unbalanced optimal transport alignments

FUGW stands for Fused Unbalanced Gromov Wasserstein [Thual et al., 2022]. It leverages optimal transport solvers from Séjourné et al., 2021 and Vayer et al., 2020, and minimizes a loss function designed to match points between two graphs based on feature similarity while preserving the underlying geometry of these graphs. In our case, features represent fMRI activity throughout a series of experiments, but they could as well integrate anatomical data such as myelin concentration, cortical depth, etc. The underlying graph geometry is that of the cortical sheet.

We assume that all individuals have an n -vertex mesh. For a subject s , we denote $F^s \in \mathbb{R}^{n,m}$ the matrix of m features per vertex, $D^s \in \mathbb{R}^{n,n}$ the matrix of distances between vertices on the cortical sheet².

Given a pair of individuals (s, t) , we seek to derive a coupling $P \in \mathbb{R}^{n,n}$ minimizing

$$L(P) \triangleq (1 - \alpha) L_W(P) + \alpha L_{GW}(P) + \rho L_U(P) + \varepsilon H(P) \quad (4.1)$$

where

$L_W(P) \triangleq \sum_{i,k} \|F_{i,\cdot}^s - F_{j,\cdot}^t\|_2^2 P_{i,k}$ is the Wasserstein loss between features,

$L_{GW}(P) \triangleq \sum_{i,j,k,l} |D_{i,j}^s - D_{k,l}^t|^2 P_{i,k} P_{j,l}$ is the Gromov-Wasserstein loss between cortical geometries,

$L_U(P) \triangleq \text{KL}(P_{\#1} \otimes P_{\#1} | \frac{1}{n} \otimes \frac{1}{n}) + \text{KL}(P_{\#2} \otimes P_{\#2} | \frac{1}{n} \otimes \frac{1}{n})$ is the marginal constraint allowing to only partly transport some areas of the cortex,

$H(P) \triangleq \text{KL}(P \otimes P | (\frac{1}{n} \otimes \frac{1}{n}) \otimes (\frac{1}{n} \otimes \frac{1}{n}))$ is the entropic regularisation term,

$\text{KL}(\cdot, \cdot)$ is the Kullback-Leibler divergence, $P_{\#1} \triangleq (\sum_j P_{i,j})_{0 \leq i < n}$ is the first marginal of P , $P_{\#2} \triangleq (\sum_i P_{i,j})_{0 \leq j < n}$ is the second marginal of P , and $\alpha \in [0, 1]$, $\rho, \varepsilon \in \mathbb{R}_+$ are the hyper-parameters of the problem.

4.1.2.2 Experiment

Baseline Since all individual data have been projected onto a common anatomical template, given a pair of individuals, one can derive the Pearson

²We compute geodesic distances between vertices using <https://github.com/the-virtual-brain/tvb-gdist>

77 **Section 4.1** Aligning human subjects with short acquisition-time fMRI training data

correlation between feature maps of the test set without functional alignment. The average correlation $b_{s,t}$ between feature maps serves as a baseline correlation for this pair of subjects.

Metric Given a pair of subjects (s, t) and a coupling $\mathbf{P}^{s,t}$ minimizing Eq. 4.1, we define $\phi_{s \rightarrow t} : X \in \mathbb{R}^{n,p} \mapsto ((\mathbf{P}^{s,t})^T X) \oslash \mathbf{P}^{s,t} \in \mathbb{R}^{n,p}$ where \oslash is the element-size division. $\phi_{s \rightarrow t}$ projects any feature map X from s onto t . We use it to project new features acquired on the source subject onto the target subject’s anatomy, and compute the correlation between projected features and the same features on the target subject. We finally compare the average correlation $c_{s,t}$ to the baseline correlation $b_{s,t}$.

Dataset In this work, we exclusively rely on the Individual Brain Charting dataset [Pinho et al., 2018]. This longitudinal study on 12 human subjects consists of a large collection of fMRI contrasts acquired for a wide variety of tasks. It also comes with movie-watching data, and other features we don’t use in this work such as retinotopic and T1-weighted maps. We derive couplings from two types of training data:

- task-based fMRI features ($m = 60$), which can be simple conditions – results of a one-sample t-test on condition maps – or contrasts – results of a t-test run on a linear combination of conditions
- movie-watching Shared Response Model (SRM) components ($m = 20$) that have been computed on the whole cohort. Naturalistic stimuli include *Raiders of the Lost Ark*, short video clips and auditory stimuli from *The Little Prince* respectively adapted from Bhattasali et al., 2019; Haxby et al., 2011; Nishimoto et al., 2011

Validation data, which we use to select a good set of hyper-parameters for Eq. 4.1, consist of a collection of 45 fMRI contrast maps acquired for a wide variety of tasks (motor, language, emotional). Test data consist of a collection of 30 fMRI contrast maps acquired for tasks related to mathematics and language.

SRM Share response models seek to find a common dictionary D of activation patterns across subjects $s \in S$ and to derive a mapping W^s that projects each individual’s features onto this common space:

$$\begin{aligned} \min_{(\mathbf{P}_s)_{s \in S}, D} \sum_{s \in S} \|\mathbf{F}^s - \mathbf{P}_s D\|_2^2 \\ \text{s.t. } \mathbf{P}_s^T \mathbf{P}_s = I \end{aligned}$$

We leverage work from Richard, Gresele, Hyvarinen et al., 2020 to compute the first m components of this representation.

4.1.3 Results

In Fig. 4.1, we show that alignments computed with SRM components using movie-watching data from *Raiders of the Lost Ark* allow to increase correlation between fMRI contrast maps for any pair of individuals of the dataset under study (panel A) and that gains happen mostly in the prefrontal and parietal lobes (panel B).

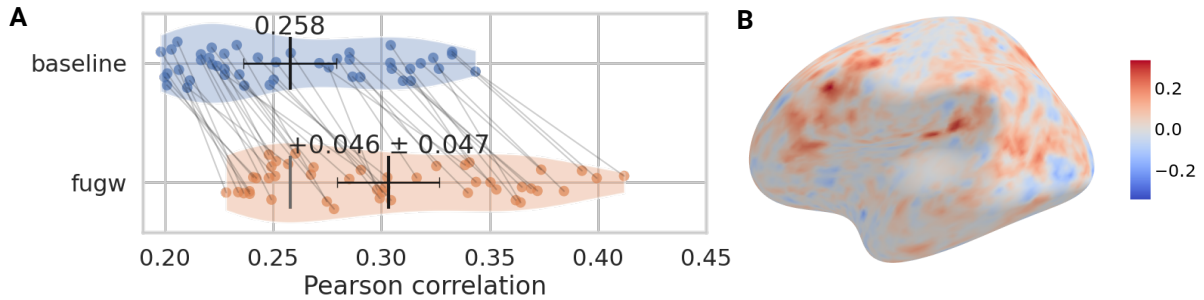


Figure 4.1: **A.** Correlation gain on the test set for the left hemisphere for each pair of individuals of the IBC dataset aligned using *Raiders of the Lost Ark* SRM components **B.** Average correlation gain per vertex for the same train and test data

In Table 4.1, we report acquisition time per individual (presence in the scanner) as well as correlation gain on the test set for a series of affordable alignment data. We recall correlation gains from Thual et al., 2022 induced by alignments derived using a training dataset consisting of all IBC fMRI contrasts except those of the validation and test sets. It seems that naturalistic stimuli are not as efficient as task fMRI contrast maps to align subjects, but that a limited number of contrast maps are enough to derive close-to-SOTA couplings.

Table 4.1: Acquisition time (AT) and correlation gain on the left hemisphere (CG) per training set (baseline correlation = 0.258)

Training set	m	Type	Acquisition time (min)	Correlation gain
All IBC (SOTA)	399	tasks	2000	0.098
Archi protocol	60	tasks	60	0.072
Bang	20	movie	8	-0.009
Clips	10	movie	100	0.009
Good Bad Evil	20	movie	180	0.039
Raiders	20	movie	115	0.045
The Little Prince	20	audio	100	0.004
All combined	90	movie + audio	503	0.050

4.1.4 Discussion

Given a cohort of individuals, aligning their respective signals is a necessity in order to detect significant activity patterns occurring in areas of the cortex

which show high inter-subject variability [Fedorenko, 2021]. However, individuals cannot be aligned using data on which one wants to infer population-level effects, hence the need for affordable data protocols that can be acquired along with protocols of interest and used for alignment.

In this work, we used multiple short-acquisition-time training datasets available in IBC. Two of them (*Raiders* and *Archi*) can bring significant correlation gains between subjects on new fMRI data. The *Archi* protocol stands out as a very efficient and affordable candidate so as to collect alignment data on large cohorts of human subjects.

In our future work, we will replicate these experiments using resting-state data. One could also select a minimalist sub-sample of tasks from IBC in order to design an even shorter protocol maximizing correlation gain.

4.2 Dependence to initialisation

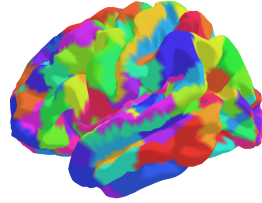
The underlying optimization problem of FUGW is not convex, therefore it is not guaranteed that the solver will converge to the same solution given different initialisations. In this section, we investigate the dependence of FUGW on the initialisation of the solver.

To this end, we first introduce a way to quickly visualise how the cortex is reorganised by the computed transport plan: given a transport plan P , one can use the associated projector ϕ defined in Equation 3.6 to project each RGB channel of the coloring of an atlas of the source participant onto the target participant. In particular, at the end of each BCD iteration, we can visualise the reorganisation of the cortex at this stage.

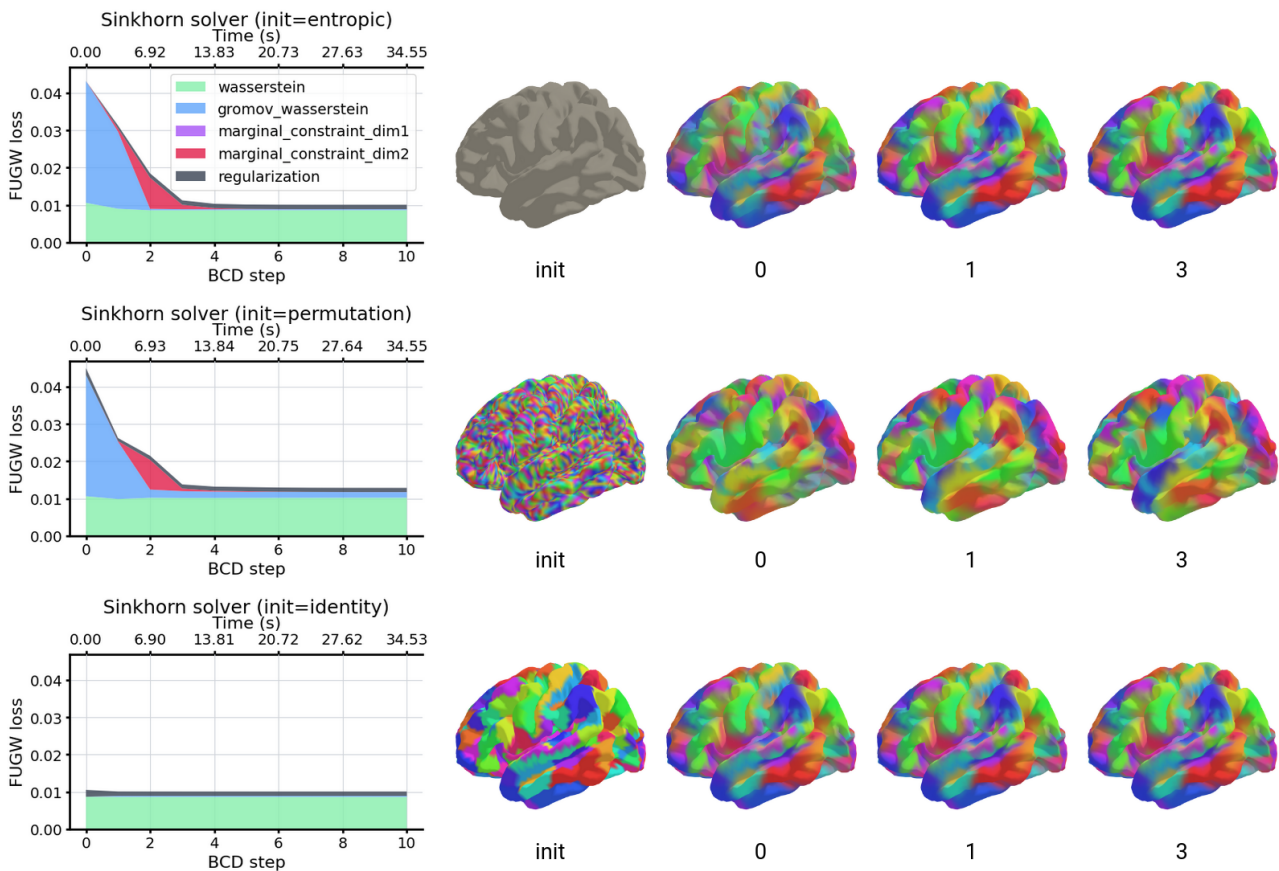
We test three different initialisation methods:

1. entropic: we initialise the transport plan as $w^s \otimes w^t$, where w^s and w^t are the weights of the source and target individuals
2. permutation: we initialise the transport plan with random values
3. identity: in the case where the data of the source and target participants lies on the same mesh (fsaverage 5 for instance), we initialise the transport plan with the identity matrix

Results are reported in Figure 4.2. Figure 4.2a shows a coloring of the MMP 1.0 atlas [Glasser et al., 2016] displayed on the left hemisphere of fsaverage 5. Figure 4.2b shows the convergence of FUGW with the three different initialisation strategies. Empirically, we find that the entropic initialisation is very robust, and that the permutation initialisation is not robust. In particular, we observe that the permutation initialisation can lead to anatomically implausible transport plans, as shown in Figure 4.2b. On the contrary, the identity initialisation leads to minimal changes in the transport plan: in practice, the entropic regu-



(a) MMP 1.0 Atlas [Glasser et al., 2016] displayed on left hemisphere of fsaverage 5



(b) Visualisation of the convergence of the FUGW loss (left column) and transported MMP 1.0 atlas from source to target at different BCD steps (right columns) for the entropic, permutation and identity initialisation strategies (rows 1, 2 and 3 respectively).

Figure 4.2: Effect of three different initialisation strategies on the convergence of FUGW. (a) Coloring of the MMP 1.0 atlas on the source individual ; we transport it to the target individual using FUGW at different stages of the training procedure. (b) We observe that the permutation initialisation can lead to anatomically implausible transport plans. The other two initialisation strategies are robust and compute similar transport maps.

larisation blurs the transport plan a little, and mild anatomical displacements are made. Note that the identity initialisation is not possible in the general case where the source and target individuals lie on different meshes.

4.3 New solvers for FUGW

In Chapter 3, we introduced a solver implementing a Block Coordinate Descent (BCD) approach to minimise a lower bound of the FUGW loss. The BCD algorithm is written in Algorithm 3.1. During each of the BCD steps, it runs many iterations of Chizat et al., 2019’s scaling algorithm - detailed in Algorithm 3.3 - approximating the solution to an unbalanced Wasserstein problem. However, optimal transport research is rich with many solvers for such optimal transport problem. Therefore, we enrich our BCD approach so that it can use different solvers than that of Chizat et al., 2019. Namely, we add two solvers coming from existing alternatives: the first one is based on a majorisation-minimisation (MM) algorithm described in [Chapel et al., 2021], the second one on an extension to the unbalanced setting of the inexact-bregman-proximal-point (IBPP) algorithm described in [Xie et al., 2020]. These additions were made in collaboration with Huy Tran.

This effort is motivated by two intuitions:

1. these two solvers have been shown to be faster than Sinkhorn’s algorithm in certain cases, and we thought they might be faster than the scaling algorithm as well
2. contrary to Sinkhorn and the scaling algorithms, the MM and IBPP algorithms allow to benefit from the sparsity of the transport plan throughout the iterations of the BCD algorithm ; we will show that this is important to scale FUGW to high-resolution data in section 4.4

Algorithm 4.1 MM algorithm from Chapel et al., 2021

Input: $K, w_s, w_t, \rho_s, \rho_t, \varepsilon$

Output: Optimal coupling P

- 1: $\lambda = \rho_s + \rho_t + \varepsilon$
 - 2: $L = (w_s^{(\rho_s + \varepsilon)/\lambda} \otimes w_t^{(\rho_t + \varepsilon)/\lambda}) \odot \exp(-K/\lambda)$
 - 3: **while** has not converged **do**
 - 4: $P \leftarrow (P^{(\rho_s + \rho_t)/\lambda} \odot L) \oslash (P_{\#1}^{\rho_s/\lambda} \otimes P_{\#2}^{\rho_t/\lambda})$
 - 5: **end while**
-

4.4 Scaling FUGW for high spatial resolutions

The work presented in this section was published in the form of a poster at OHBM 2023. It aims at showing that it is possible to solve FUGW problems even when the number of vertices of the source and target brains is high. This would prove useful to align high-resolution data (collected at 7T for instance)

Algorithm 4.2 IBPP algorithm from [Xie et al., 2020](#)

Input: $K, w^s, w^t, \rho, \varepsilon, k$
Output: Optimal coupling P

- 1: $L = \exp(-K/\rho)$
- 2: $P = w^s \otimes w^t$
- 3: **while** has not converged **do**
- 4: $R \leftarrow P \odot L$
- 5: **for** $i \in \{1, \dots, k\}$ **do**
- 6: $u \leftarrow (R(v \odot w^t))^{\frac{-\rho}{\rho+\varepsilon}}$
- 7: $v \leftarrow (R^T(u \odot w^s))^{\frac{-\rho}{\rho+\varepsilon}}$
- 8: **end for**
- 9: $P \leftarrow (u \otimes v) \odot R$
- 10: **end while**

as well as volumetric images, whose number of voxels typically exceeds 10242 (the number of vertices of fsaverage 5).

The main issue stems from the fact that the size of transport plans grows quadratically with the number of vertices of the source and target brains. Let us denote n and p the number of vertices of the source and target cortices respectively, and P a transport plan mapping them. P is a matrix of size $n \times p$.

In all previous experiments, we worked on the fsaverage 5 mesh, which has 10242 vertices per hemisphere. Consequently, P was a matrix of size $10\,242 \times 10\,242 \approx 104$ M coefficients. Empirically, a matrix of such size storing float32 coefficients takes approximately 400MB in memory.

Consequently, computing a transport plan between two fsaverage 7 meshes would result in a matrix of size $163\,842 \times 163\,842 \approx 26$ G coefficients, taking approximately 102GB in memory. It is not possible to fit such a matrix in memory on a single GPU, and computations would be very slow on a CPU. Moreover, in the framework developed in Chapter 3, many such matrices need to be stored in memory to optimise for the FUGW loss.

4.4.1 A coarse-to-fine approach to scale FUGW

We propose to use a coarse-to-fine approach to scale FUGW to high-resolution data. The idea is to first compute a dense transport plan between two low-resolution samplings from the high-resolution meshes, and then use it to compute a sparse transport plan between the two high-resolution meshes.

Below are the detailed steps to compute P a sparse transport plan between two high-resolution meshes, which we also report formally in Algorithm 4.3:

1. Randomly select a subset of vertices of the source and target meshes
2. Fit a dense transport plan \tilde{P} between these subsets
3. Define the sparse matrix P_0 by computing the following sparsity mask:
 For any pair of vertices (i, k) from the source and target subsets respec-

tively maximally matched in \tilde{P} , allow all vertices within $r = 5\text{mm}$ of i to be mapped to vertices within $r = 5\text{mm}$ of k

4. Fit P using the same procedure as described in Chapter 3, starting from P_0

Algorithm 4.3 Coarse-to-fine LB-FUGW for high-resolution data

Input: $\mathcal{X}^s, \mathcal{X}^t, r, \alpha, \rho, \varepsilon$

Output: Optimal coupling P

- 1: $X^s, X^t \leftarrow \text{LMDS}(D^s), \text{LMDS}(D^t)$
 - 2: $S^s, S^t \leftarrow \text{SampleUniformly}(X^s), \text{SampleUniformly}(X^t)$
 - 3: $\tilde{P} \leftarrow \text{LB-FUGW}(\mathcal{X}_{S^s}^s, \mathcal{X}_{S^t}^t, \alpha, \rho, \varepsilon)$
 - 4: $P_0 \leftarrow \text{SparseTransportPlan}(\tilde{P}, X^s, X^t, r)$
 - 5: $P \leftarrow \text{LB-FUGW}_{\text{sparse}}((F^s, X^s, w^s), (F^t, X^t, w^t), P_0, \alpha, \rho, \varepsilon)$
-

Figure 4.3 illustrates the difference between dense and sparse mappings, as well as the steps described above. Besides, we typically use a radius of 5-10 millimeters to build the aforementioned neighbourhoods. Empirically, this results in sparse transport plans comprising 100 M non-null coefficients and taking approximately 1GB in memory. We could fit the whole procedure on a single Tesla V100-DGXS-32GB GPU, and it would take approximately 10 minutes to compute P .

Note that other approaches to scale FUGW to high-resolution data could be considered. In particular, it is possible to compute low-rank approximations of the transport plan. Indeed, if the computed transport plan is low-rank, it can be stored in a much more memory-efficient way. [Forrow et al., 2018](#) introduces a natural way to compute low-rank solutions to the Wasserstein problem by computing two plans: one transporting the source space to a low-dimensional intermediate space, and the other transporting the intermediate space to the target space. Recent papers have extended this approach to unbalanced and fused settings [[Scetbon, Cuturi and Peyré, 2021](#); [Scetbon, Peyré and Cuturi, 2021](#); [Scetbon et al., 2023](#)]. These approaches add factorisation constraints forcing computed solutions to be low-rank. One nice property of these approaches is that, when applied to brain data, dimensions of the intermediate space can be interpreted as clusters of vertices/voxels, which can be useful to interpret the computed transport plan as well as to build atlases of the brain.

However, we advocate that our coarse-to-fine approach allows computing high-rank solutions, which may prove useful to align high-resolution brain data. As an example, mixing together functional and anatomical connectivity data acquired at high resolution might result in data whose real rank is high.

4.4.2 Adapting FUGW to use sparse matrices

At first, it is unclear whether imposing a sparsity mask on P will allow one to fit the entire FUGW procedure on a single GPU. Indeed, other matrices of the

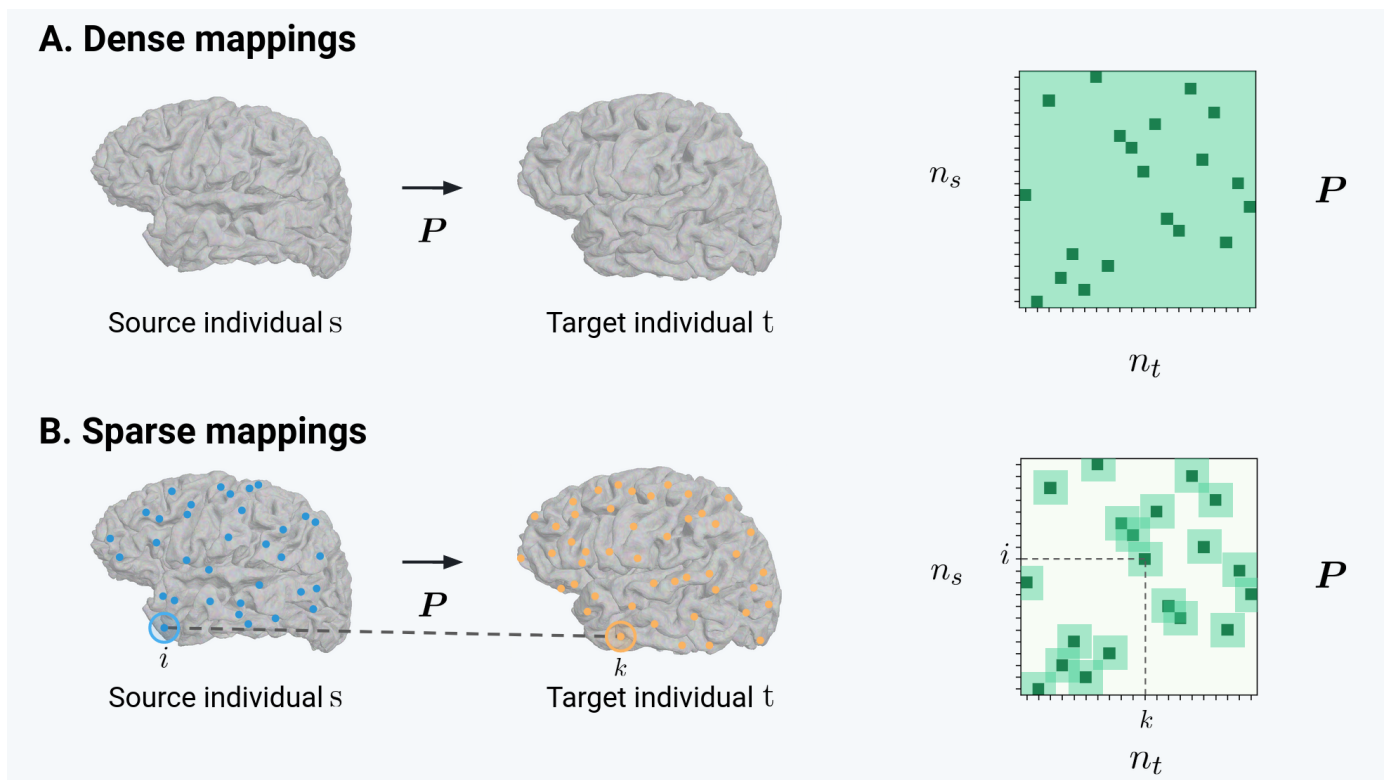


Figure 4.3: Storing sparse matrices allows scaling FUGW to higher brain resolutions This figure illustrates the difference between dense and sparse mappings. **A.** All coefficients (light and dark green) of P are stored, but only a handful of them are not null (dark green). **B.** For high-resolution meshes, instead of storing all coefficients, we only store a selection of them (light and dark green) in a sparse matrix. Non-selected coefficients are set to 0. The selection is determined by mapping random subsets of vertices of the source and target individuals. These subsets are represented as blue (resp. orange) points for the source (resp. target) participant. Vertices i and k from the source and target random subsets respectively were maximally matched, and thus neighbouring vertices of i are allowed to be mapped to neighbouring vertices of k . We typically choose a radius of 5-10 millimeters to build these neighbourhoods.

algorithm might end up being too large to fit in memory as n and p increase. Let us list them and design strategies to fit them in memory.

In particular, the lower-bound approximation algorithm detailed in Algorithm 3.1 requires to compute cost matrices $K \in \mathbb{R}^{n \cdot p}$. Let us notice an important difference between Sinkhorn and the scaling algorithm on one hand, and the MM and IBPP algorithms on the other hand: the former use all coefficients of K to update the dual variables u and v , whereas the latter only use coefficients within the sparsity mask of P . From a computational perspective, it stems from the fact that the MM and IBPP algorithms compute updates of P that consist in element-wise products of P and K (actually, a point-wise function of K). Therefore, one only needs to store coefficients of K that are within the sparsity mask of P .

Consequently, each part of the equation defining K reported in Algorithm 3.2, namely C the matrix of pairwise distances between functional signatures, G the geometry cost tensor, and the unbalancing and regularisation terms, can be stored using sparse matrices with the sparsity mask of P .

Finally, Q the other transport plan fitted jointly with P in Algorithm 3.1 is also of size $n \times p$. We propose using the same sparsity mask for P to fit Q .

4.4.2.1 Adapting the LB-FUGW algorithm

In theory, the previous section shows that the cost matrix K need only be evaluated on the support of the sparse transport plan P . In practice, our efficient function for computing the cost matrix detailed in Algorithm 3.2 needs to be adapted. In particular, it uses the matrices of geodesic distances between the source and target meshes vertices, D^s and D^t respectively. These matrices are of size $n \times n$ and $p \times p$, respectively, which will not fit in memory on a single GPU when n and p are high.

In the following paragraphs, we (1) propose to compute low-rank approximations of these matrices which can be stored in a memory-efficient way, and (2) show that the approximated matrix of geodesic distances can be factorised, which allows us to adapt Algorithm 3.2 to remain computationally efficient.

Low-rank approximation of the matrix of geodesic distances We propose to use a low-rank approximation \tilde{D} of the matrices of geodesic distances, i.e. to find $X^s \in \mathbb{R}^{n \times k}$ where k is small and such that $\tilde{D}^s \triangleq \left(\|X_i^s - X_j^s\|_2^2 \right)_{i,j}$ and $\tilde{D}^s \approx D^s$ (respectively $X^t \in \mathbb{R}^{p \times k}$ such that $\tilde{D}^t \triangleq \left(\|X_i^t - X_j^t\|_2^2 \right)_{i,j}$ and $\tilde{D}^t \approx D^t$). Finding such vectors, often referred to as *embeddings*, is the goal of Multidimensional Scaling (MDS). Many algorithms have been proposed to solve this problem, and we choose to use the Landmark Multidimensional Scaling (LMDS) algorithm described in Platt, 2005 to compute X^s and X^t . Like other algorithms performing this task, it is based on the Nyström approximation of the eigenvectors and eigenvalues of the distance matrices. Subsequently, we

contribute a Python implementation of the LMDS algorithm, which we add to the `fugw` package.

Factorization of the low-rank matrix of geodesic distances Let us denote $\tilde{D} \triangleq \left(\|\mathbf{X}_i - \mathbf{X}_j\|_2^2 \right)_{0 \leq i, j < n} \in \mathbb{R}^{n \times n}$. We seek to derive D_1 and D_2 such that $\tilde{D} = D_1 D_2^T$. Let us remark that

$$\begin{aligned} D_1 &= (\text{norm}(\mathbf{X}, 1)^T, \mathbb{1}_n^T, -\sqrt{2}\mathbf{X}) && \in \mathbb{R}^{n, k+2} \\ D_2 &= (\mathbb{1}_n^T, \text{norm}(\mathbf{X}, 1)^T, \sqrt{2}\mathbf{X}) && \in \mathbb{R}^{n, k+2} \end{aligned}$$

are solutions to this problem, where $\text{norm}(\mathbf{X}, 1) \in \mathbb{R}^n$ is the vector containing the squared L2-norms of each row of \mathbf{X} and $\mathbb{1}_n \in \mathbb{R}^n$ is the vector of ones.

Adapting the computation of the cost matrix and couplings With the previously described low-rank embeddings and factorisation, we can adapt the LB-FUGW algorithm to use sparse, high-resolution matrices.

First, we need to update the computation of the cost matrix K in Algorithm 3.2. Let us denote D_1^s, D_2^s the factorisation of D^s obtained from the previous section, and D_1^t, D_2^t the factorisation of D^t . The term $D^s P (D^t)^T$ can be efficiently derived by computing matrix multiplications a proper order. Indeed, the following association runs without memory issues using the sparse matrix multiplication method `torch.sparse.mm()` from PyTorch:

$$D_1^s \left[\left[(D_2^s)^T \left[P D_2^t \right] \right] (D_1^t)^T \right]$$

The other terms can be adapted in a similar way.

Secondly, we need to update the algorithms deriving the transport plan P . The MM algorithm detailed in Algorithm 4.1 can easily be adapted to use sparse matrices.

4.4.3 Sampling high-resolution meshes uniformly

Leveraging the low-rank embeddings introduced in the previous section, we derive a method to sample points on a high-resolution mesh such that they are uniformly distributed on the surface. Let $\mathbf{X} \in \mathbb{R}^{n \times k}$ be the matrix of vertices of the high-resolution mesh. We simply run the Ward algorithm on \mathbf{X} to derive m approximately equal-sized clusters of vertices, and then sample points uniformly within each cluster.

4.5 Precision mapping of human extrastriate visual cortex with high-resolution fMRI inter-subject alignment

The work presented in this section was published in the form of a short article and poster at CCN 2023. It directly uses the coarse-to-fine approach to scale FUGW to high-resolution data developed in the previous section, and aims at showing that it is possible to align high-resolution fMRI data using FUGW.

4.5.1 Introduction

First introduced in the early 1990s, functional Magnetic Resonance Imaging (fMRI) has since become a major tool to non-invasively study human cognition [Bandettini et al., 1992; Ogawa et al., 1992]. While initial spatial resolution was around 3mm - see for example Vinckier et al., 2007 for a study on the ventral visual pathway - increasingly high magnetic-field strengths [De Martino et al., 2018; Dumoulin et al., 2018] as well as denoising techniques [Vizioli et al., 2021] have recently made it possible to acquire images at sub-millimeter resolution.

Increase in spatial resolution in fMRI commands for fine-grained analysis tools Such increase in spatial resolution offers new exciting possibilities of in-vivo mesoscale imaging, for example the unraveling of columnar-like organizations of the human cortex [Schneider et al., 2019]. However, it also requires new methods for the jointed statistical inference of functional data across different individuals.

Indeed, at common spatial resolutions, a conventional way of conducting group-level analysis is to project each individual on a common anatomical template and spatially smoothing BOLD data to increase sensitivity, which results in a considerable loss of spatial details. At increasingly high (e.g., sub-millimeter) resolutions, because individual variability in anatomy becomes prominent, such group-level analysis can no longer be conducted, as illustrated in Zhan et al., 2023. Instead, careful inspection of individual-level data is required, with many manual steps in a time-consuming task at risk of insufficient reproducibility between researchers.

4.5.2 Methods

Dataset We rely on data published in Zhan et al., 2023, in which authors acquire 7T fMRI on 21 French-English bilingual subjects to look for word-specific sub-regions of the VWFA. We restricted our analysis to the 7 early-bilingual subjects who did not have bias in either languages. Anatomical data were acquired using a 0.65mm isotropic MP2RAGE sequence, resampled to 0.6mm isotropic and transformed into MNI space. Individual surfaces were subsequently obtained using FreeSurfer v6 [Fischl, 2012]. Functional data were acquired with

a 2D gradient-echo echo-planar imaging (EPI) sequence with a voxel size of 1.2mm isotropic. After standard and minimal preprocessing, data were transformed into MNI space and kept at 1.2mm isotropic. No smoothing of functional data occurred at this stage. Functional data consisted of one functional localizer run (276 volumes) containing object categories, and three main fMRI runs (399 volumes each) containing letter strings increasingly similar to real words. See [Zhan et al., 2023](#) for more detailed description of the experiment and preprocessing steps.

In this study, we use maps of t-statistics on volumetric data and project them on individual meshes using the python package Nilearn³.

Baseline conventional alignment Using Freesurfer, we can map individual maps from individual surfaces onto a common anatomical template (fsaverage 7). At this step, we can smooth individual data on the surface. Finally, we run a t-test in each vertex to evaluate which vertices are significantly activated at the group level. Empirically, we increased the smoothing value millimeter by millimeter until we reached a value returning significant activations at the group level (3mm smoothing was needed).

Aligning individuals using anatomical and functional data We use FUGW v0.1⁴, an alignment technique developed in [Thual et al., 2022](#) which now scales to high mesh resolutions. For a given pair of individuals, this method derives a mapping which maximizes functional correlation while minimizing anatomical displacement. This framework is equipped with a distance, which enables one to evaluate similarity between two individuals with regard to these two objectives.

Using ten contrasts in the functional localizer data (including single-category contrasts like faces, body, tools, houses), we derive an alignment for each of the 7 subjects on a reference subject. Computed mappings are then used to align each main experimental condition. The reference subject on whom all others were aligned was chosen to minimize the FUGW distance to all other subjects. Hyper-parameter values were left to the package's defaults.

After FUGW alignment, we run a group analysis similar to that of the baseline.

4.5.3 Results

In figure 4.4, we show that, compared to conventional group-level analysis, our method successfully recovers meaningful activation clusters at the group level. In particular, we highlight cortical regions significantly activated ($p \leq 0.001$, uncorrected) in the main experimental conditions. These regions correspond to word-sensitive areas identified in the functional localizer. While baseline

³<https://nilearn.github.io>

⁴<https://alexisthual.github.io/fugw>

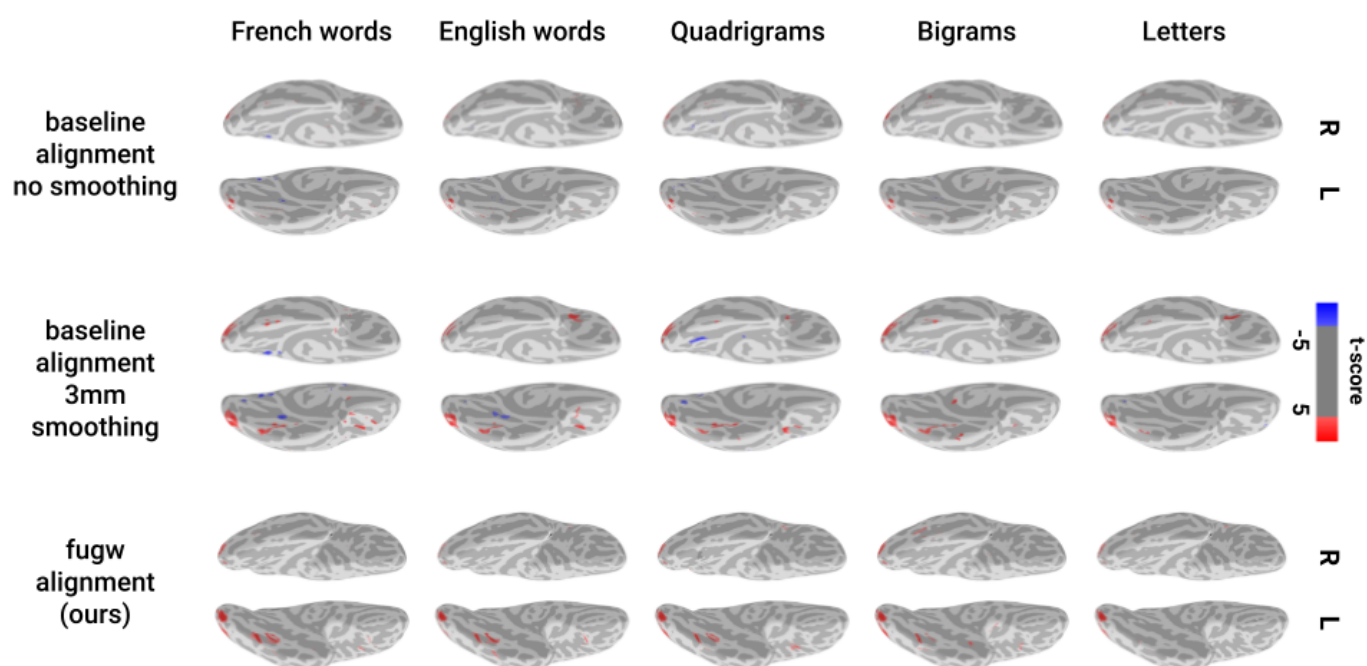


Figure 4.4: Group analysis comparison shows that FUGW recovers precise activation clusters at population level. In these ventral views of t-score maps plotted with Nilearn, subjects are shown letter strings (letters, bigrams, quadrigrams, real words) increasingly similar to real words. Conventional group analysis on un-smoothed data (left) fails to recover meaningful activation clusters at the group level. Conventional analysis on 3mm smoothed data (middle) recovers some activation patches as well as hard-to-interpret "negative blobs". Data aligned with FUGW on functional localizers with multiple object categories (right) exhibit multiple cortical patches in the VWFA. These patches are increasingly anterior as stimuli get more similar to real words.

approaches involving Freesurfer and fsaverage fail to align individual signal, our approach recovers interpretable and fine-grained clusters, showing left-predominant activation in the ventral occipito-temporal cortex (VOTC) consistent with previous findings and theoretical predictions [Dehaene et al., 2005].

Similarly to Zhan et al., 2023, we see that the VWFA comprises multiple patches of cortex rather than a single continuous occipito-temporal area as was initially described in Vinckier et al., 2007. Additionally, we provide evidence supporting the idea that the ventral visual stream is organized along a posterior-to-anterior gradient, with increased sensitivity to stimuli similar to real words in the most anterior cortical patches.

4.5.4 Discussion

Recent advances in fMRI tools providing highly resolved sub-millimeter functional data has opened exciting new perspectives for the precise mesoscopic study of brain function. Yet, current analysis techniques do not offer the proper framework for meaningful group-level analysis that keep the fine-grained details. We here show that FUGW can successfully help tackle these difficulties. In particular, we illustrate how using a data-driven study-specific precise anatomical template (in our case, a reference subject within our cohort), can help recover meaningful activation clusters at the group level.

Limitations of this work include projecting volumetric t-maps on surfaces: a more precise approach would require recomputing these statistics on surfaces directly.

Following steps will include analysis of the entire dataset of the aforementioned study, including 14 additional French-English bilingual subjects (7 English-dominant and 7 French-dominant) and 10 English-Chinese bilingual subjects. We plan to precisely map - at the group level - the cortical patches of the ventral visual cortex specialised for each language. We also plan to provide additional evidence for the posterior-to-anterior gradient of the ventral stream with more anterior regions showing higher sensitivity to real words than more posterior ones.

4.6 Comparing the convergence speed of three solvers for the FUGW problem

Speed is an important aspect of FUGW, since users may want to align a large number of pairs of individuals in parallel. We therefore test all three algorithms with different numbers of iterations during each BCD step, and compare their convergence speed in Figure 4.5. We find that both MM and IBPP – introduced in Section 4.3 – can run a much higher number of iterations than Scaling during each BCD step, but that their convergence is slower. Eventually, we find that,

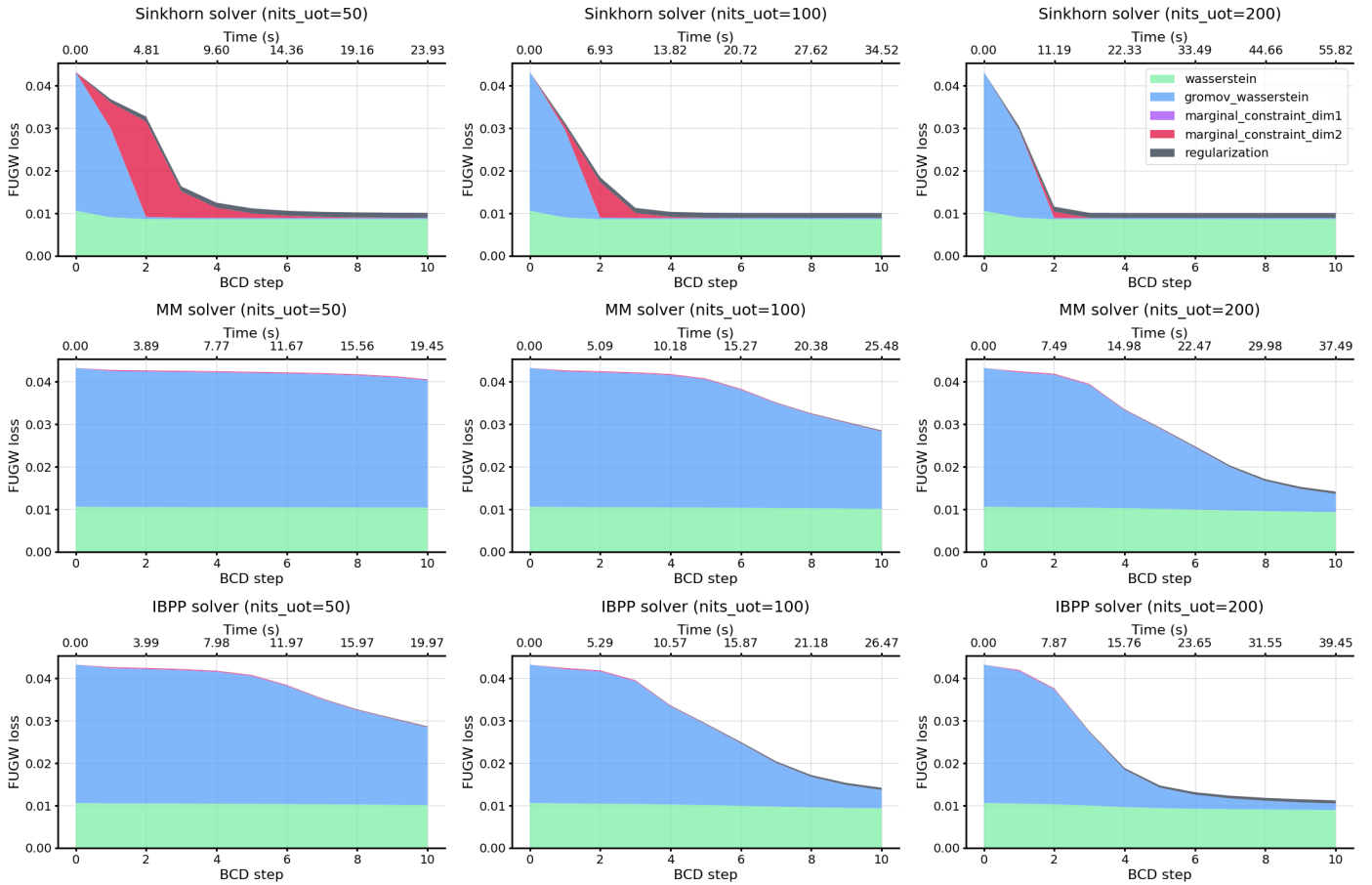


Figure 4.5: Convergence speed comparison of three solvers for the FUGW problem when fitting using 23 IBC contrast maps of the Archi Standard protocol. BCD steps are shown on the bottom x-axis, and computation time is on the top x-axis. We track the value of the FUGW loss during training on the y-axis. Iterations of the MM and IBPP algorithms (second and third rows respectively) are faster than those of the Scaling algorithm (first row). However, their convergence is slower. With our data of interest, we see that 100 iterations of the Scaling algorithm ran 5 to 6 times (one time for each BCD iteration) are enough to reach convergence in about 20 seconds on a single Tesla V100-DGXS-32GB.

with our implementations and data, Scaling is the fastest and MM is the slowest, and that 6 BCD iterations of 100 Scaling iterations each are enough to reach convergence in about 20 seconds on a single Tesla V100-DGXS-32GB.

Note that, for this experiment, features used to fit the transport plan are fMRI contrast maps of the left hemisphere taken from two participants of the IBC dataset (sub-04 and sub-07). Namely, we take all 23 contrast maps from the Archi Standard protocol⁵. However, we evaluate the speed of convergence on different contrast maps in the same participants. Namely, we take all 119 contrast maps from the MVIS, MVEB, Lec1, Lec2, Audi, MSCE, Visu, and Moto protocols⁵. All contrast maps are represented on fsaverage 5 (10242 vertices per hemisphere).

Lastly, in order to check that our conclusion holds when the number of contrast maps used to fit the transport plan is higher, we run the same experiment with 100 and 220 contrast maps by first adding the Archi and HCP protocols and then the RSVPLanguage, Preference, MathLanguage, FaceBody and Emotion protocols. We report all results for these two setups in Figures 10.1 and 10.2. However, changing the mesh resolution - i.e. the number of vertices - would probably require tuning the number of BCD and Scaling iterations again.

4.7 Replacing the Kullback-Leibler divergence with the squared L2-norm in the FUGW problem

Empirically, we observed that solutions computed by the LB-FUGW algorithm introduced in Chapter 3 showed very little absolute difference in transported mass between vertices. We hypothesised that this was due to using the Kullback-Leibler divergence in the marginal constraints and regularisation term. Therefore, we decided to implement a new series of solvers for the FUGW problem using the squared L2-norm instead of the Kullback-Leibler divergence, as the former may allow less smooth distributions of mass than the latter.

In the following paragraphs, we denote as $L_{\theta}^{\text{bi}}(P, Q)$ the loss function of the FUGW problem when using the squared L2-norm instead of the Kullback-Leibler divergence. We show in Theorem 4.7.2 that, similarly to the case of the KL divergence, minimising $L_{\theta}^{\text{bi}}(P, Q)$ with respect to P while fixing Q is equivalent to solving a regularised unbalanced Wasserstein optimal transport problem. Before that, we prove Lemma 4.7.1, which is used extensively in the proof of Theorem 4.7.2. In the following paragraphs, we write $\|\cdot\|_2$ as $\|\cdot\|$ for simplicity.

Lemma 4.7.1. *Let $a, b, c, d \in \mathbb{R}^n$. There exist $\mu \in \mathbb{R}^n$ and $\tau, \kappa \in \mathbb{R}_+$ which can be expressed as functions of b, c and d only, such that*

$$\|a \otimes b - c \otimes d\|^2 = \tau \|a - \mu\|^2 + \kappa$$

⁵All IBC protocols are described in <https://individual-brain-charting.github.io/docs/tasks.html>

93 Section 4.7 Replacing the Kullback-Leibler divergence with the squared L2-norm in the FUGW problem

and the following values for τ, μ, κ are solutions:

$$\tau = \|\mathbf{b}\|^2 \quad ; \quad \mu = \frac{\langle \mathbf{b}, \mathbf{d} \rangle}{\|\mathbf{b}\|^2} \mathbf{c} \quad ; \quad \kappa = \|\mathbf{c}\|^2 \left(\|\mathbf{d}\|^2 - \frac{\langle \mathbf{b}, \mathbf{d} \rangle^2}{\|\mathbf{b}\|^2} \right)$$

Proof.

$$\begin{aligned} \|\mathbf{a} \otimes \mathbf{b} - \mathbf{c} \otimes \mathbf{d}\|_2^2 &\triangleq \langle \mathbf{a} \otimes \mathbf{b} - \mathbf{c} \otimes \mathbf{d}, \mathbf{a} \otimes \mathbf{b} - \mathbf{c} \otimes \mathbf{d} \rangle \\ &= \langle \mathbf{a} \otimes \mathbf{b}, \mathbf{a} \otimes \mathbf{b} \rangle - 2\langle \mathbf{a} \otimes \mathbf{b}, \mathbf{c} \otimes \mathbf{d} \rangle + \langle \mathbf{c} \otimes \mathbf{d}, \mathbf{c} \otimes \mathbf{d} \rangle \\ &= \|\mathbf{a}\|^2 \|\mathbf{b}\|^2 - 2\langle \mathbf{a}, \mathbf{c} \rangle \langle \mathbf{b}, \mathbf{d} \rangle + \|\mathbf{c}\|^2 \|\mathbf{d}\|^2 \\ &= \|\mathbf{b}\|^2 \left(\|\mathbf{a}\|^2 - 2\langle \mathbf{a}, \frac{\langle \mathbf{b}, \mathbf{d} \rangle}{\|\mathbf{b}\|^2} \mathbf{c} \rangle \right) + \|\mathbf{c}\|^2 \|\mathbf{d}\|^2 \\ &= \|\mathbf{b}\|^2 \left(\left\| \mathbf{a} - \frac{\langle \mathbf{b}, \mathbf{d} \rangle}{\|\mathbf{b}\|^2} \mathbf{c} \right\|^2 - \left\| \frac{\langle \mathbf{b}, \mathbf{d} \rangle}{\|\mathbf{b}\|^2} \mathbf{c} \right\|^2 \right) + \|\mathbf{c}\|^2 \|\mathbf{d}\|^2 \\ &= \|\mathbf{b}\|^2 \left\| \mathbf{a} - \frac{\langle \mathbf{b}, \mathbf{d} \rangle}{\|\mathbf{b}\|^2} \mathbf{c} \right\|^2 + \|\mathbf{c}\|^2 \left(\|\mathbf{d}\|^2 - \frac{\langle \mathbf{b}, \mathbf{d} \rangle^2}{\|\mathbf{b}\|^2} \right) \end{aligned}$$

□

Moreover, Lemma 4.7.1 can be generalised to the case of tensors of order d by noticing that the proof only uses scalar products, and that the scalar product of two tensors of order d can be written as the scalar product of their vectorised forms. Let us now state and prove Theorem 4.7.2.

Theorem 4.7.2. For a fixed Q , there exist $\mathbf{K}, \Gamma \in \mathbb{R}^{n,p}, \mu \in \mathbb{R}^n, \nu \in \mathbb{R}^p$ and $\tau_1, \tau_2, \lambda \in \mathbb{R}_+$ such that the optimal $P \in \arg \min_{P \geq 0} L_\theta^{\text{bi}}(P, Q)$ is the solution of

$$\min_{P \geq 0} \langle \mathbf{K}, P \rangle + \tau_1 \|P_{\#1} - \mu\|^2 + \tau_2 \|P_{\#2} - \nu\|^2 + \lambda \|P - \Gamma\|^2$$

Proof. Let us recall that

$$L_\theta^{\text{bi}}(P, Q) \triangleq (1 - \alpha) L_W(P, Q) + \alpha L_{\text{GW}}(P, Q) + \rho L_U(P, Q) + \varepsilon E(P, Q)$$

where

$$\begin{aligned} L_U(P, Q) &\triangleq \|P_{\#1} \otimes Q_{\#1} - \mathbf{w}^s \otimes \mathbf{w}^s\|^2 + \|P_{\#2} \otimes Q_{\#2} - \mathbf{w}^t \otimes \mathbf{w}^t\|^2 \\ E(P, Q) &\triangleq \|P \otimes Q - (\mathbf{w}^s \otimes \mathbf{w}^t) \otimes (\mathbf{w}^s \otimes \mathbf{w}^t)\|^2 \end{aligned}$$

Let us fix the value of Q . We are going to identify $\mathbf{K}, \mu, \nu, \Gamma, \tau_1, \tau_2, \lambda$ successively, by rewriting the terms of $L_\theta(P, Q)$ in elements which depend on P , and elements which do not. The latter can be discarded because, as we optimise with respect to P , they do not affect the solution.

First, let us derive K :

$$\begin{aligned} (1 - \alpha) L_W(P, Q) + \alpha L_{GW}(P, Q) &\triangleq (1 - \alpha) \langle C, \frac{P + Q}{2} \rangle + \alpha \langle G, P \otimes Q \rangle \\ &= \langle \frac{(1 - \alpha)}{2} C + \alpha G \otimes Q, P \rangle + \langle \frac{1 - \alpha}{2} C, Q \rangle \\ &= \langle K, P \rangle + \kappa_1 \end{aligned}$$

where $K \triangleq \frac{(1 - \alpha)}{2} C + \alpha G \otimes Q$ and κ_1 does not depend on P .

Secondly, let us use Lemma 4.7.1 three times to rewrite all elements of $L_U(P, Q)$ and $E(P, Q)$:

$$\begin{aligned} \|P_{\#1} \otimes Q_{\#1} - w^s \otimes w^s\|^2 &= \tau_1 \|P_{\#1} - \mu\|^2 + \kappa_2 \\ \|P_{\#2} \otimes Q_{\#2} - w^t \otimes w^t\|^2 &= \tau_2 \|P_{\#2} - \nu\|^2 + \kappa_3 \\ \|P \otimes Q - (w^s \otimes w^t) \otimes (w^s \otimes w^t)\|^2 &= \lambda \|P - \Gamma\|^2 + \kappa_4 \end{aligned}$$

where

$$\begin{aligned} \tau_1 &\triangleq \|Q_{\#1}\|^2, & \mu &\triangleq \frac{\langle Q_{\#1}, w^s \rangle}{\|Q_{\#1}\|^2} w^s \\ \tau_2 &\triangleq \|Q_{\#2}\|^2, & \nu &\triangleq \frac{\langle Q_{\#2}, w^t \rangle}{\|Q_{\#2}\|^2} w^t \\ \lambda &\triangleq \|Q\|^2, & \Gamma &\triangleq \frac{\langle Q, w^s \otimes w^t \rangle}{\|Q\|^2} w^s \otimes w^t \end{aligned}$$

and $\kappa_2, \kappa_3, \kappa_4$ do not depend on P .

Eventually, we have showed that $L_\theta^{\text{bi}}(P, Q)$ can be rewritten as

$$L_\theta^{\text{bi}}(P, Q) = \langle K, P \rangle + \tau_1 \|P_{\#1} - \mu\|^2 + \tau_2 \|P_{\#2} - \nu\|^2 + \lambda \|P - \Gamma\|^2 + \kappa_5$$

where κ_5 does not depend on P . Therefore minimising $L_\theta^{\text{bi}}(P, Q)$ with respect to P while fixing Q is equivalent to solving an L2-regularised unbalanced Wasserstein optimal transport problem. \square

The optimisation problem arising from Theorem 4.7.2 is already known and can be solved using the MM algorithm from Chapel et al., 2021 recalled in Algorithm 4.1.

Chapter 5

Preliminary work on inter-species comparisons

In Chapter 3, we introduced FUGW, a method to compute whole-brain mappings between individuals with minimal anatomical priors. We showed that it could successfully align fMRI data acquired in different human participants. In Chapter 4, we showed that it could be extended to work on high-resolution data and that it was possible to change the loss such that it better models the expansion/shrinkage of cortical areas.

In this Chapter, we leverage previously developed extensions of FUGW to align cortical data coming from different species. We aim to push forward recent efforts made in cross-species comparisons [Eichert et al., 2020; Mantini et al., 2012; Xu et al., 2020]. In the present study, we collect and analyse a functional MRI dataset comprising human participants and macaques stimulated with the same visual, auditory, and tactile stimuli. In particular, we compute alignments between all possible pairs of human participants and macaques. These alignments try to map vertices from the two cortical surfaces to maximize functional similarity while limiting anatomical displacements. Our mappings also model how easy it is for a voxel to map from one subject to another under the previously mentioned constraints.

We qualitatively show that our method correctly maps functional areas known to be common across the two species (visual, motor, auditory). Our quantitative analyses are coherent with the existing literature, showing that primary areas are proportionally bigger in macaque subjects than in human subjects. More importantly, our data-driven study also reveals that parietal and pre-frontal areas involved in language processing are proportionally much bigger in human subjects than in macaque ones, thus beginning to pinpoint the better-developed circuits in human subjects and their functional roles. This work lays new foundations for comparing cortices between species but should still be considered preliminary. In particular, this approach should dramatically benefit from richer

comparative fMRI data sets, which could also help assess the relevance of optimal transport approaches to such comparisons.

5.1 Methodology and data collection

5.1.1 Dataset

We thank Wim Vanduffel, and especially Marcelo Armendariz, for collecting the data and providing the following paragraphs explaining how the multi-modal dataset was acquired in human participants and macaques.

5.1.1.1 Subjects

Three rhesus monkeys (*Macaca mulatta*, two female, 4-6 years old, 5-8 kg) and twenty healthy human volunteers (12 female, 19-35 years old) participated in the study. Human volunteers were informed about the experimental procedures and signed a written informed consent form. Animal care and experimental procedures met the Belgian and European guidelines and were approved by the ethical committee of the KU Leuven Medical School. Animals were born in captivity and were pair- or group-housed (two to five animals per group; cage size at least 16-32 m³) with cage enrichment (toys, foraging devices), outside views and natural day-night cycles (throughout the year supplemented with an artificial 12/12hr light/dark cycle) at the primate facility of the KU Leuven Medical School.

They were daily fed with standard primate chow supplemented with bread, nuts, raisins, prunes and fruits. The animals received their water supply either during the experiments, or in the cages before and after the experiments. Monkeys had previous experience performing behavioural tasks and were prepared for fMRI sessions. Prior to scanning, monkeys were trained daily (2-5 weeks) to perform a passive fixation task while in a sphinx position with their head rigidly fixed in a plastic primate chair. Details concerning head-post surgery and training procedures have been previously described [Vanduffel et al., 2001]. During the experimental period, access to water was restricted but animals were allowed to drink until fully satiated during the daily training and scanning sessions.

5.1.1.2 Stimuli

The stimulus set comprised three modalities: visual, auditory and tactile. Visual stimuli consisted of ten classes of achromatic images - monkey and human faces, monkey and human bodies (excluding the head), four-legged mammals, birds, man-made objects (matched either to the monkey or to the human bodies), fruits/vegetables, and body-like sculptures. With a total of 100 pictures, each class consisted of 10 images which were previously used in the fMRI study of [Popivanov et al., 2012]. The vertical and horizontal extent of the images

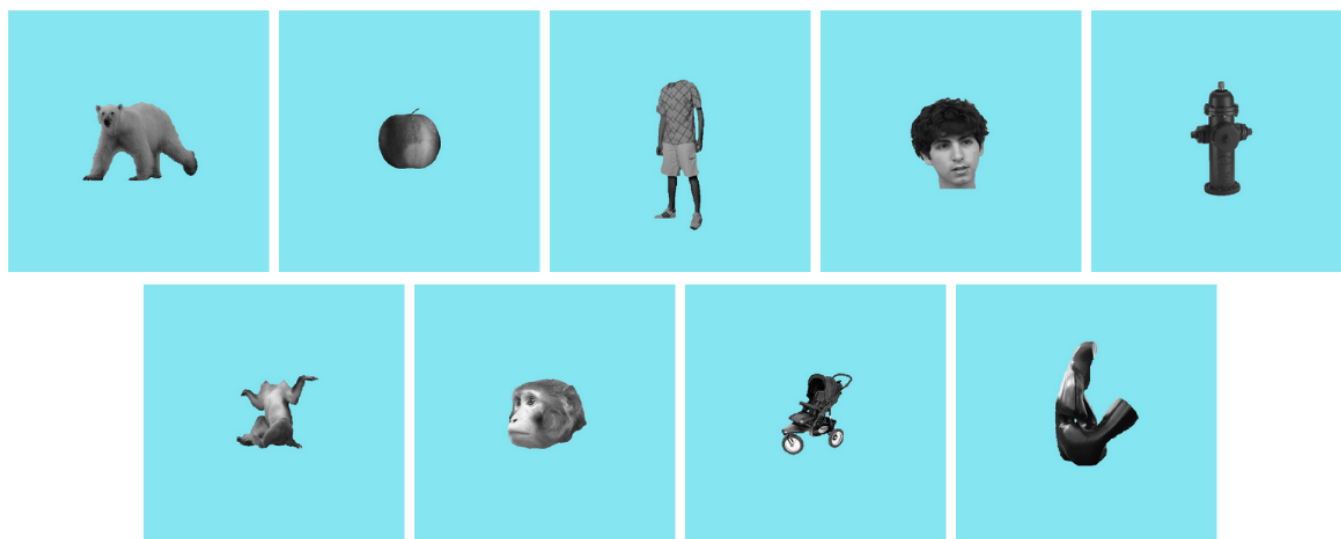


Figure 5.1: Monkey and human subjects are shown images belonging to different categories From top to bottom, left to right, are images taken from the animal, fruit, human body, human face, human object, monkey body, monkey face, monkey object, sculpture categories respectively.

ranged from 5° to 10° of visual angle. The images were embedded into a pink noise background having the same mean luminance as the images and which filled the entire display ($30^\circ \times 40^\circ$ of visual angle). Auditory stimuli consisted of recordings of various natural sounds that included six categories - human speech, human non-speech sounds (e.g., baby cry, laughter, coughing), monkey calls, other animal sounds (e.g., lion, horse), tool sounds and musical instruments (e.g., keys, scissors, piano, flute), and scenes from nature (e.g., rain, wind, thunder). Each category included 10 different sounds which were previously used in [Moerel et al., 2012], with the exception of the monkey calls that were added for this study. Tactile stimulation consisted of air puffs delivered to three different locations on the lower left and right sides of the face of the subjects: beneath the lower lip, above the upper lip and the cheek area. Air puffs were delivered with an intensity of 0.5 bars at a distance of 5 mm from the face. To control for any potential sound produced by the puffs, two extra air puffs were located adjacent to the face, but pointing in the opposite direction.

5.1.1.3 Experimental setup and design

Human volunteers laid in a supine position and watched the screen through a mirror tilted 45° towards a translucent screen onto which the visual stimuli were projected at a frame rate of 60 Hz. Participants had to fixate passively on a red dot (0.25°) presented in the centre of the screen. Eye position was monitored at 120 Hz using a pupil-corneal reflection tracking system (Iscan). Monkeys were placed within the bore of the magnet in sphinx position inside a plastic primate chair using a physical head restraint.

Images were projected (Barco 6300 LCD projector) on a translucent screen located at a distance of 57 cm from the monkey. As in humans, subjects had to passively fixate while eye position was monitored. To encourage monkeys to maintain fixation and remain quiet, liquid reward was delivered through a plastic tube located just inside their mouths. MR-compatible headphones with ear-cup pad were used in both species to deliver the acoustic stimuli at approximately 75 dB SPL, and to shield the ears from environmental noise [Joly et al., 2012].

For tactile stimulation we built a computer-controlled MR-compatible pneumatic system with articulated plastic arms that we used to automatically and systematically deliver air puffs to different locations of the face surface of humans and monkeys inside the scanner bore [Huang and Sereno, 2007]. Importantly, we also introduced a control air puff, which was not directed to the face or body of the subjects. Before each scanning session, a contrast agent, monocrySTALLINE iron oxide nanoparticle (MION), was injected into the femoral/saphenous vein (6-11 mg/kg) to improve the contrast-to-noise ratio [Leite et al., 2002; Popivanov et al., 2012]. As we aimed to compare evoked responses to the same sensory stimulation, we followed identical experimental protocols for every subject and species, particularly with regard to the order and timing of the events. We generated eight different sequences of stimuli where the three modalities were randomly and separately presented in an event-related manner to the awake and fixating subjects. A red dot located in the center of the screen and superimposed on an achromatic pink noise background was permanently present on the screen and served as control. To ensure a balanced presentation of the three modalities along the eight sequences, stimuli were repeated accordingly: three times for images (3×100); five times for sounds (5×60); and one hundred times for tactile stimuli (100×3). In addition, controls for images and sounds (same control for both: background and silence, 150 repetitions) and for tactile stimuli (background and control air puffs, 150 repetitions) were randomly presented. As a result, each sequence contained a random selection of 150 stimuli or controls. Our design was conditioned by the fact that sounds needed to be presented during silent periods (no scanner noise) to be clearly audible [Erb et al., 2019]. Thus, to preserve uniformity across modalities, all the stimuli were presented during 1000 ms in the center of a silent gap (1200 ms) between scan acquisitions.

5.1.1.4 fMRI data acquisition

Monkey data were acquired with a 3T MR Siemens Trio scanner with an AC88-insert head gradient. Functional images were collected using a gradient-echo T2*-weighted echo-planar imaging sequence (repetition time (TR) = 2600 ms, acquisition time (TA) = 1400 ms, echo time (TE) = 15 ms, 37 slices, voxel size = 1 mm isotropic, flip angle (FA) = 75°) and MION contrast agent. Before every TA, a stimulus was presented in a silent gap of 1200 ms. As a result, each trial (2600-ms, TR) consisted of: 1400-ms (TA) + 100-ms silence + 1000-ms condi-

tion (visual, audio, tactile or control) + 100-ms silence. Four dummy scans were added at the beginning and the end of each sequence during which the fixation dot and the visual background were presented. Each monkey participated in 5-7 scanning sessions where the eight stimuli sequences were repeated until the animal stopped fixating.

Monkeys were scanned with a custom-built, 8-channel, implanted phased-array receive coil (Janssens et al., 2012) and a saddle-shaped radial transmit-only surface coil. To provide an anatomical reference for the functional scans, high-resolution T1-weighted images were acquired for each monkey during a separate session under ketamine-xylazine anesthesia using a single radial transmit-receive surface coil and a MPRAGE sequence (TR = 2200 ms, TE = 4.05 ms, FA = 13°, 208 slices, voxel size = 0.4 mm isotropic). During the session, 12-15 whole-brain volumes were obtained and averaged to improve signal-to-noise ratio. In humans, fMRI acquisition was performed with a 3T MR Philips Achieva scanner using a 32-channel head coil. Functional images were obtained using T2*-weighted echo-planar images (TR = 2600 ms, TA = 1400 ms, TE = 30 ms, 44 slices, voxel size = 2.3 × 2.3 × 2.5 mm, FA = 75°) with blood oxygen level-dependent contrast (BOLD). Participants underwent one scanning session where the eight different sequences were presented. For each subject, a high-resolution T1-weighted image was acquired using an MPRAGE sequence (TR = 8.1 ms, TE = 3.7 ms, voxel size = 0.9 × 0.9 × 1 mm).

5.1.1.5 Image preprocessing

We preprocessed the fMRI data using Matlab (MathWorks), SPM12 software package (Wellcome Trust Centre for Neuroimaging¹) and JIP².

We only analyzed runs where subjects performed more than 95% of fixation within a 1.5° × 1.5° fixation window (160 runs for humans and 224 runs for monkeys). Temporal preprocessing was applied to correct for slice-dependent time shifts, head motion and linear trends. Then, we spatially warped the monkey and human functional data to F99 and MNI atlas spaces, respectively. Next, we high-pass filtered the data and removed non-neuronal signal fluctuations by regressing out the mean signal from a ventricular region of interest and a region centered in the white matter. Data were spatially smoothed with a Gaussian kernel at 3 and 6 mm FWHM for monkeys and humans, respectively. To allow inter-species comparison, we accounted for differences in hemodynamic response functions (HRFs) by convolving the monkey and human fMRI timecourses with a canonical HRF (BOLD from SPM and MION [Erb et al., 2019; Popivanov et al., 2012]) from the other species [Mantini et al., 2012]. Finally, for each subject, runs (150 volumes each) were concatenated in ascending sequence order (from 1 to 8) and z-scored, yielding 20 individual human datasets. In the case of monkeys, for which several repetitions for each sequence were acquired per subject, we

¹<https://www.fil.ion.ucl.ac.uk/spm>

²<https://www.nitrc.org/projects/jip>

split the data and generated 20 separated monkey concatenations. As a result, we obtained 20 human and 20 monkey individual datasets, in which voxel timecourses were composed of 1200 volumes (150 x 8 sequences). Finally, we also generated a group dataset for each species by averaging the timecourses in corresponding voxels across the 20 individual datasets. This procedure allowed us to maximize the relative contribution of stimulus-evoked responses exceeding spontaneous activity in our analysis.

Moreover, all macaque subjects are anatomically aligned to the MEBRAINS template. We project their volumetric data to a mesh extracted from MEBRAINS using Freesurfer. Analogously, we use Freesurfer to compute the individual anatomical mesh of each human subject, and project their data to their respective individual anatomies. We chose to use a common anatomical template for macaque subjects because their anatomy is less variable across individuals than it is for humans, while using a template can ease group analyses.

5.1.1.6 General Linear Model

For each subject and each run, we fit a general linear model that regresses task-related activations of the brain, as well as potential confounding factors (like movement). In particular, we regress activations for each of the conditions presented in the Stimuli subsection. We then fit a fixed-effect model using all these fitted GLMs.

Initial conditions are detailed in the Stimuli subsection. We build compound conditions called contrasts from these initial conditions to gain statistical power, and compute statistical maps for these contrast maps. They are detailed in Table 5.1. Figure 5.2 illustrates what such contrast maps look like for a subset of subjects. Interestingly, Figure 5.2 shows that monkey subjects react to monkey-made sounds but not to human-made sounds, and vice-versa. Consequently, we introduce two conditions which depend on the subject's species: same-species-speech vs other, which for a human (resp. macaque) subject will human speech and voice sounds (resp. macaque speech sounds) as positive conditions, and all other audio conditions as negative ones in our model.

5.1.2 Alignments setup and computation

We make use of this unique inter-species multi-modal dataset to test extensions made to FUGW in Chapter 4.

5.1.2.1 Using L2 penalization instead of Kullback-Leibler divergence for marginal constraints

The role of marginal constraints is to model the fact that some cortical areas exist in a given subject and not in the other, or that they have expanded / shrunk

Contrast map	Positive conditions	Negative conditions
face vs other	human faces, monkey faces	animals, birds, fruits, human bodies, human objects, monkey bodies, monkey objects, sculptures
bodies vs other	human bodies, monkey bodies	all other visual conditions
bodies vs other non faces	human bodies, monkey bodies	all other visual conditions except human faces and monkey faces
animate vs inanimate	human faces, human bodies, monkey faces, monkey bodies, animals, birds	all other visual conditions
monkey speech vs other	monkey sounds	animal, nature, human speech, human voice, tools sounds
(speech+voice) vs other	human speech and voice sounds	all other sounds
audio vs control	all audio conditions	control condition
tactile vs audio	all tactile conditions	all audio conditions
tactile vs control	all tactile conditions	control condition
visual vs audio	all visual conditions	all audio conditions
visual vs control	all visual conditions	control condition
visual vs tactile	all visual conditions	all tactile conditions

Table 5.1: Explicit description of contrast maps from regressed conditions

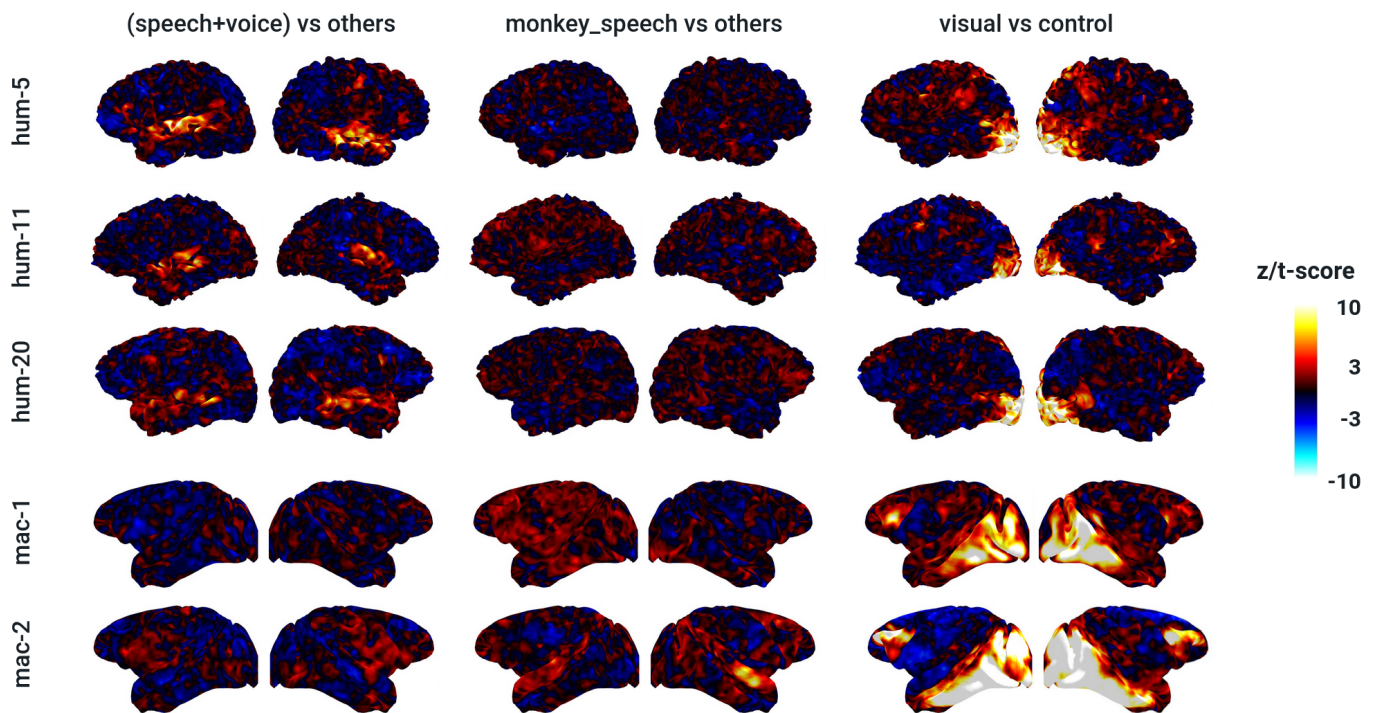


Figure 5.2: Two contrast maps for five subjects of the inter-species dataset
This figure shows t-statistics derived from z-score maps fitted in each fMRI run. These z-score maps are obtained by contrasting clear and unclear human speech conditions against all other auditory conditions (column 1), the clear "monkey speech" condition against all other auditory conditions (column 2) and all visual conditions against a control condition (column 3). We show these maps for 3 human subjects (first 3 rows) and 2 macaque subjects (last two rows).

from one subject to the other. We draw such conclusions from how much mass was transported from / to a vertex relative to other vertices.

It is common practice in Optimal Transport to formulate the marginal constraints and regularization term using the Kullback-Liebler divergence. Empirically, we noticed that using the Kullback-Leibler divergence flattens a lot the distribution of transported masses. In short, in the data regime underlying this study, it is hard to interpret masses differences when using the KL divergence.

We reasoned that an L2 penalization would allow for more salient and meaningful transported mass discrepancies between vertices, which motivated our additions in section 4.7.

5.1.2.2 Using sparse mappings to transport high-resolution meshes

As explained in section 4.4, one major impediment of FUGW is that computed matrices scale quadratically with the anatomical resolution of the data considered. In particular, the number of vertices in a macaque cortical mesh is typically 100k, while the number of vertices of fsaverage 7 for human participants is 160k. Consequently, computing a dense mapping between these two subjects would require storing a 100k x 160k matrix, which would take approximately 100 GB of disk space. We opt to use the coarse-to-fine approach designed in section 4.4 to compute sparse mappings between human and macaque subjects.

In order to compute this sparse mapping, we first generate two random subsets of points from the source and target subjects respectively. We size these subsets so that a dense mapping between them will fit on memory, and such that the number of vertices in each subject is close to that of experiments made in [Thual et al., 2022](#). By doing so, we can use our knowledge of the best hyper-parameters to use in this case.

The second step consists in building a sparsity mask for mapping which will be fit during the next step. For this, we leverage information contained in the mapping computed at the previous step: for each vertex i of the source subject, we find the maximally matched vertex k in the target subject, and enable all combinations of neighbours of i and neighbours of k to be stored in the mapping. Neighbours of i are defined as vertices that are within radius distance r of i on the cortical sheet, for which we take the mid-thickness surface, defined as the surface between the white and pial surfaces.

Finally, we solve a new FUGW problem on a sparse mapping matrix whose sparsity mask was pre-computed at the previous step. Typically, taking subsets of 10k points from meshes of 160k points and setting our radius to 7 millimeters yields mappings which use approximately 1.5 GB of disk storage.

5.1.2.3 Computing alignments

Using the coarse-to-fine approach described in section 4.4, for each possible pair of human and monkey subjects, we compute two optimal transport alignments (one for left hemispheres, one for right hemispheres) using contrast maps described in section 5.1.1. This results in $20 \times 3 \times 2 = 120$ mappings in total, each of which takes approximately 10 minutes to compute on a Tesla V100-DGXS-32GB and takes approximately 1 GB of disk space. These mappings were derived in parallel on a Slurm cluster.

5.2 Results: Functional Homologies and Differences

We explore computed mappings from a qualitative and quantitative perspective. Our qualitative analyses mostly consist in visually exploring computed mappings in order to make sure that computed alignments make sense. Quantitative results aim at show that the location of cortical areas with more or less transported mass is coherent with existing literature.

5.2.1 Qualitative results

In order to explore our mappings precisely, we designed and developed brain-cockpit, a web application which we give more details about in section 8.1.2. In particular, brain-cockpit can allow exploring computed mappings the following way: users can click on a specific vertex of a subject who was previously aligned to another subject, and see the probability map of associated vertices in the second subject. Figure 5.3 illustrates this approach. These qualitative checks allow us to verify that computed mappings make sense: in particular, they allow us to control that primary areas are correctly mapped across individuals.

However, while brain-cockpit allows one to precisely check what our mappings have computed, a more comprehensive view of derived alignments can be obtained by using them to transport an atlas from the source subject to the target subject (or vice-versa).

More precisely, we colour each parcel of a given atlas using a swatch containing colours with equivalent energy (ie, considering each colour is a 3-dimensional vector, all these vectors have the same norm). We then transport each of the 3 colour channels from source to target, allowing us to generate a transported version of the source atlas on the target anatomy. Figure 5.4 illustrates this method, and shows that primary areas are correctly mapped across human and macaque individuals.

5.2.2 Quantitative results

On top of associating vertices of the source and target subjects, computed mappings also model "how much" a given vertex can be mapped between sub-

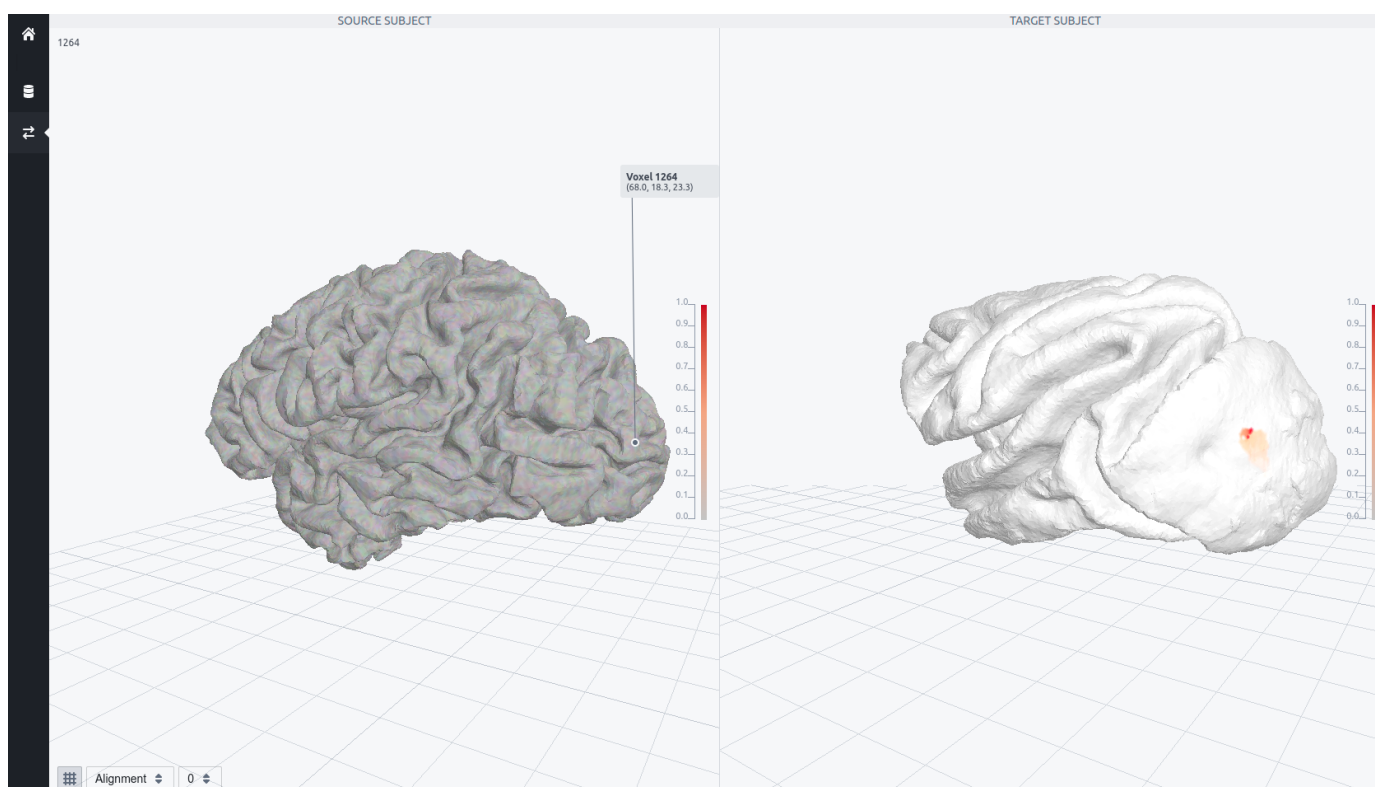


Figure 5.3: Brain-cockpit, a web-application to explore computed human / macaque mappings A vertex is selected in the occipital lobe of a human subject (left) for which we computed a mapping to a macaque subject. The probability map of associated vertices in this macaque subject is displayed (right). Brain-cockpit allows selecting any vertex from the source / target subject.

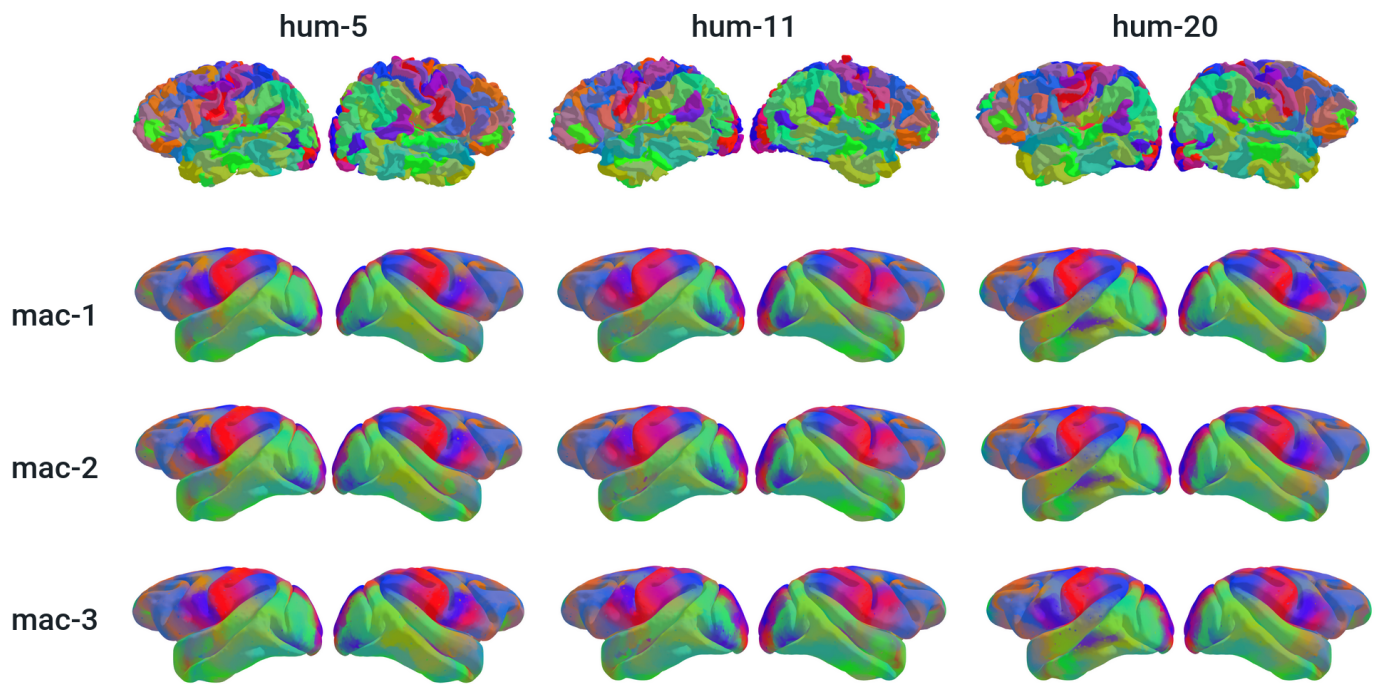


Figure 5.4: Transporting the Glasser atlas from human subjects to monkey subjects This figure shows the Glasser atlas [Glasser et al., 2016] for 3 human subjects of the dataset (top row), and projections of these atlas to each of the 3 macaque subjects of our dataset (following rows). In particular, we observe that anatomy is generally well preserved when mapping human subjects to macaque subjects.

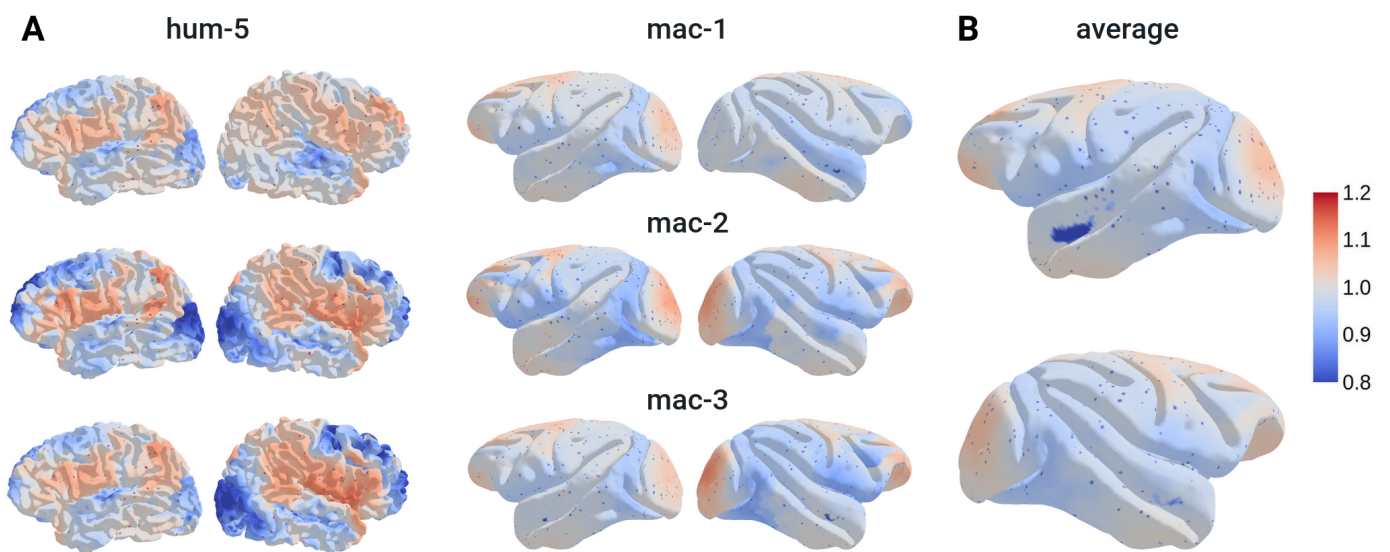


Figure 5.5: Transported mass averaged across macaque subjects on the ME-BRAINS template Representation of the relative transported mass in each vertex of the cortical sheet. Red voxels gained mass, blue ones lost mass. It can be interpreted the following way: blue areas are relatively smaller (i.e. occupy a smaller proportion of the cortical sheet) in one subject than in the other. On the other hand, red parts are relatively larger. In all images, one can observe out-of-distribution voxels: these are artefacts from approximating the anatomical geodesic distance with low-rank methods. A. Mass transported from hum-1 (left column) onto each of the 3 macaque subjects (right column). B. Transported mass in each macaque vertex, averaged across macaque subjects.

jects. This value is denoted as the mass, and corresponds to the norm of a given row (or column) of the mapping. One can compute and plot the mass in each vertex of the source and target subjects. Figure 5.5 shows this mass distribution on the cortical surface.

In particular, relative mass gives us insights about which cortical areas expanded or shrunk across subjects. Figure 5.5.B shows that the visual cortex occupies a relatively larger place in the macaque cortex than its counterpart in the human brain. Conversely, the temporal lobe and prefrontal areas occupy a relatively smaller proportion of the macaque cortex.

5.3 Conclusion

Our software tool works quite efficiently, thus allowing us to compute tens of whole-brain human-macaques cortical alignments of functional brain activity in a short time. In this respect, the present work sets a new standard for inter-species alignments, which have been primarily based on anatomy alone [Eichert et al., 2020], or on regions of interest [Neubert et al., 2014; Xu et al., 2020], or on local patches of cortex rather than the whole brain [Mantini et al., 2012].

However, this is a work in progress, and the present inter-species alignments should be taken with a grain of salt because they are only as good as the fMRI paradigm that led to the activation data. The present dataset is unique in that it comprises stimuli from three different modalities (visual, auditory, and tactile). It thus anchors the computed alignments into appropriate sensory regions, as well as comprising representative stimuli that make the human and macaque brain unique (language and communication signals). However, it is still lacking in fMRI signatures that would be unique to specific brain regions, such as motion (area MT) or social stimuli [Sliwa and Freiwald, 2017].

Chapter 6

Evaluating inter-subject alignment with visual decoding tasks

6.1 Introduction

Decoding brain activity The generative capabilities of deep learning have recently unlocked decoding mental representations from brain activity. Originally restricted to linear models [Harrison and Tong, 2009; Haynes and Rees, 2006; Mitchell et al., 2004], the decoding of brain activity can now be carried out with deep learning techniques. In particular, using functional Magnetic Resonance Imaging (fMRI) signals, significant progress has been made in the decoding of images [Z. Chen, Qing, Xiang et al., 2023; Ferrante, Ozcelik et al., 2023; Gu et al., 2023; Mai and Zhang, 2023; Ozcelik and VanRullen, 2023; Scotti et al., 2023; Takagi and Nishimoto, 2023], speech [Tang et al., 2023], and videos [Z. Chen, Qing and Zhou, 2023; Kupersmidt et al., 2022; Lahner et al., 2023; Phillips et al., 2022; Wang et al., 2022; Wen et al., 2018].

The bottleneck of inter-subject variability A core issue is that brain organization is highly variable across participants, which makes it challenging to train a single model on multiple participants using fMRI data. Therefore, with few noteworthy exceptions [Haxby et al., 2020; Ho et al., 2023], studies typically train a brain decoder on a single participant at a time. With this constraint in mind, major effort has been put towards building fMRI datasets collecting a lot of data in a limited number of participants [Allen et al., 2022; LeBel et al., 2023; Pinho et al., 2018; Wen et al., 2017]. Nonetheless, the necessity to train and test models on a single participant constitutes a major impediment to using notoriously data-hungry deep learning approaches. In addition, generalization to new individuals is essential to the validation of discoveries.

Functional alignment Several methods can align the functional organization – on top of the anatomy – of multiple brains, and thus offer a potential solution to inter-individual variability: differentiable warps of the cortical surface [Robinson et al., 2014], rotations between brain voxels in the functional space [Haxby et al., 2011], shared response models [P.-H. Chen et al., 2015; Richard, Gresele, Hyvärinen et al., 2020], permutations of voxels minimizing an optimal transport cost [Bazeille et al., 2019; Thual et al., 2022], or combinations of these approaches [Feilong et al., 2022]. More recently, several studies rely on deep learning models trained in a self-supervised fashion to build an embedding of brain activity, in the hope that it could be meaningful across participants [Z. Chen, Qing, Xiang et al., 2023; Thomas et al., 2022]. However, to this day, it is not clear which of these methods offers the best performance and generalization capabilities [Bazeille et al., 2021].

Approach It is currently unknown whether any of the aforementioned methods improve the decoding of naturalistic stimuli such as videos, and how such hypothetical gain would vary with the amount of fMRI recording available in a given a participant. To address this issue, we leverage fMRI recordings of training participants to boost the decoding of videos and static images in a single left-out participant, as illustrated in Figure 6.1. This requires fitting two models: an alignment model and a brain decoder. The alignment aims at making brain responses of a left-out participant most similar to those of a reference participant. Here, we leverage optimal transport to compute this transformation using functional and anatomical data from both participants. The brain decoder – which we will refer to as the *decoder* – consists of a linear regression trained to predict the latent representations of movie frames or static images from the corresponding BOLD signals or beta coefficients. We evaluate video and image decoding in different setups. In particular, we assess (1) whether decoders generalize to participants on which they were not trained, (2) whether training a decoder on data from multiple participants improves performance and (3) the extent to which functional alignment improves the aforementioned setups.

Contributions We first confirm the feasibility of decoding, from 3 Tesla (3T) fMRI, the semantics of videos watched by the participants [Wen et al., 2017]. We verify that this approach also performs well for the decoding of static images from 7T fMRI data [Allen et al., 2022]. Our study makes three main novel contributions:

1. Compared to the baseline, functional alignment across participants boosts visual semantics decoding performance in left-out participants, especially when the latter have a limited amount of data
2. Training a decoder on multiple functionally aligned participants yields a model with improved performance compared to training one model per participant, but anatomical alignment does not

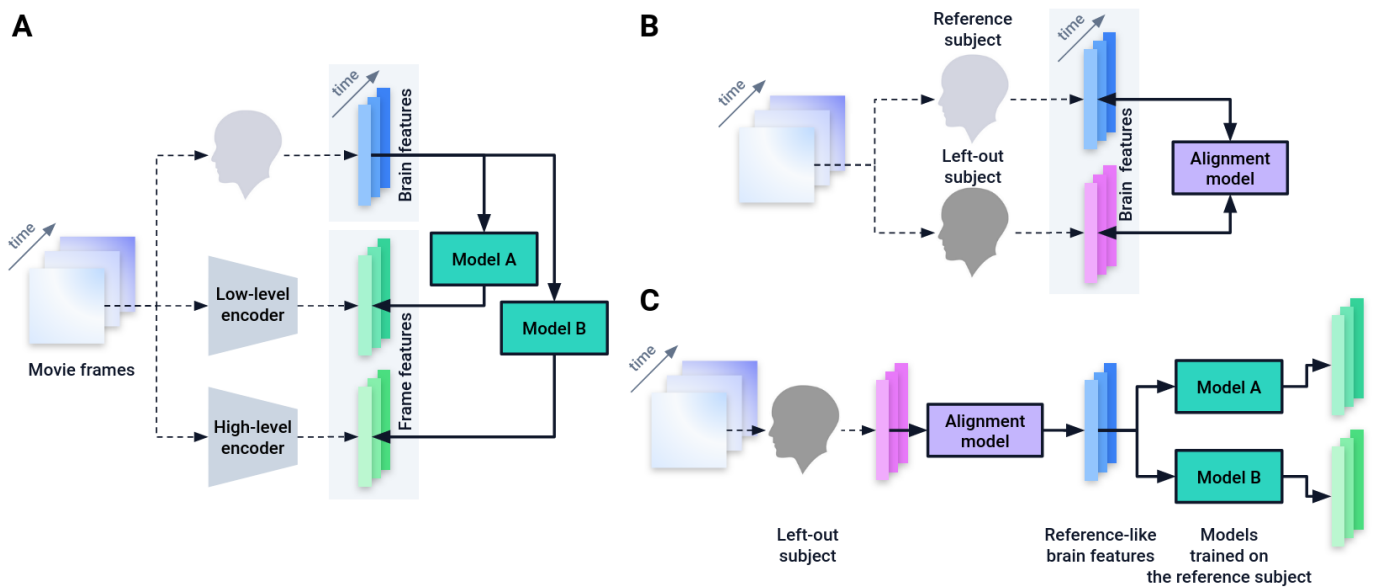


Figure 6.1: General outline of video decoding from BOLD fMRI signal in left-out participants

A. For each frame associated with a brain volume, one computes its low- and high-level latent representations using pre-trained encoders. Then, brain decoders (green) can then be fitted to map brain features onto each of these latent representations. **B.** The BOLD signal acquired from two participants watching the same movie can be used to derive an alignment model (purple) that maps voxels from the two participants based on functional similarity. **C.** Then, this alignment model can be used to transform brain features of the left-out participant into the brain features that match those of the reference participant. In particular, this allows one to use decoders that have been trained on a lot of data coming from a reference participant, and apply them on a left-out participant for whom less data was collected.

3. The resulting alignments, computed from movie-watching data, are anatomically coherent

6.2 Methods

Our goal is to decode visual stimuli seen by individuals from their brain activity. To this end, we train a linear model to predict latent representations – shortened as *latents* – of these visual stimuli from BOLD fMRI signals recorded in participants watching naturalistic videos.

In the data under study, brains are typically imaged at a rate of one scan every 2 seconds. During this period, a participant sees 60 video frames on average, or a static image for the case of [Allen et al., 2022](#). For simplicity, we consider the restricted problem of decoding only the first video frame seen by participants at each brain scan. Formally, regardless of the dataset, for a given participant, let $\mathbf{X} \in \mathbb{R}^{n,v}$ be the BOLD response collected in v voxels over n brain scans and $\mathbf{Y} \in \mathbb{R}^{n,m}$ the m -dimensional latent representation of each selected video frame.

6.2.1 Brain alignment

Anatomical alignment As a baseline, we consider the alignment method implemented in Freesurfer [[Fischl, 2012](#)], which relies on anatomical information to project each participant onto a surface template of the cortex (in our case *fsaverage5*). Consequently, brain data from all participants lie on a mesh of size $v = 10\,242$ vertices per hemisphere.

Functional alignment On top of the aforementioned anatomical alignment, we apply a recent method from [Thual et al., 2022](#) denoted as Fused Unbalanced Gromov-Wasserstein (FUGW) ¹. As illustrated in Figure 6.1.B, this method consists in using functional data to train an alignment that transforms brain responses of a given left-out participant into the brain responses of a reference participant. This approach can be seen as a soft permutation of voxels ² of the left-out participant which maximizes the functional similarity to voxels of the reference participant.

Formally, for a left-out participant, let $\mathbf{D}^{\text{out}} \in \mathbb{R}^{v,v}$ be the matrix of anatomical distances between vertices on the cortex, and $\mathbf{w}^{\text{out}} \in \mathbb{R}_+^v$ a probability distribution on vertices. \mathbf{w}^{out} can be interpreted as the relative importance of vertices; without prior knowledge, we use the uniform distribution. Reciprocally, we define \mathbf{D}^{ref} and \mathbf{w}^{ref} for a reference participant. Note that, in the general case, v can be different from one participant to the other, although we simplify notations here.

¹<https://alexisthual.github.io/fugw>

²We use the words *voxel* (volumetric pixel) or *vertex* (point on a mesh) indifferently.

We derive a transport plan $P \in \mathbb{R}^{v,v}$ to match the vertices of the two participants based on functional similarity, while preserving anatomical organisation. For this, we simultaneously optimize multiple constraints, formulated in the loss function $\mathcal{L}_\Theta(P)$ described in Equation 6.1:

$$\mathcal{L}_\Theta(P) \triangleq (1 - \alpha) \mathcal{L}_W(P) + \alpha \mathcal{L}_{GW}(P) + \rho \mathcal{L}_U(P) + \varepsilon H(P) \quad (6.1)$$

Gromov Wasserstein loss
Wasserstein loss
Marginal constraints
Regularization

Each component of the loss is expressed as follows:

- $\mathcal{L}_W(P) \triangleq \sum_{0 \leq i, j < v} \|X_i^{\text{out}} - X_j^{\text{ref}}\|_2^2 P_{i,j}$
- $\mathcal{L}_{GW}(P) \triangleq \sum_{0 \leq i, k, j, l < v} |D_{i,k}^{\text{out}} - D_{j,l}^{\text{ref}}|^2 P_{i,j} P_{k,l}$
- $\mathcal{L}_U(P) \triangleq \text{KL}(P_{\#1} \otimes P_{\#1} | w^{\text{out}} \otimes w^{\text{out}}) + \text{KL}(P_{\#2} \otimes P_{\#2} | w^{\text{ref}} \otimes w^{\text{ref}})$
- $H(P) \triangleq \text{KL}(P \otimes P | (w^{\text{out}} \otimes w^{\text{ref}}) \otimes (w^{\text{out}} \otimes w^{\text{ref}}))$

Here, $\text{KL}(\cdot, \cdot)$ denotes the Kullback-Leibler divergence, $P_{\#1} \triangleq (\sum_j P_{i,j})_{0 \leq i < n}$ is the first marginal of P , $P_{\#2} \triangleq (\sum_i P_{i,j})_{0 \leq j < n}$ is the second marginal of P , $\alpha \in [0, 1]$, $\rho \in \mathbb{R}_+$ are the hyper-parameters setting the relative importance of each constraint, and $\Theta \triangleq (X^{\text{out}}, X^{\text{ref}}, D^{\text{out}}, D^{\text{ref}}, \alpha, \rho, \varepsilon)$.

Following [Thual et al., 2022](#), we minimize $\mathcal{L}_\Theta(P)$ with 10 iterations of a block coordinate descent algorithm [[Séjourné et al., 2021](#)], each running 1 000 Sinkhorn iterations [[Cuturi, 2013](#)]. Subsequently, we define $\phi_{\text{out} \rightarrow \text{ref}}: X \mapsto (P^T X^T) \oslash P_{\#2} \in \mathbb{R}^{n,v}$ where \oslash is the element-wise division, a function that transports any matrix of brain features from the left-out participant to the reference participant. To simplify notations, for any X defined on the left-out participant, we define $X^{\text{out} \rightarrow \text{ref}} \triangleq \phi_{\text{out} \rightarrow \text{ref}}(X)$.

Hyper-parameters selection for functional alignment We use default parameters shipped with version 0.1.0 of FUGW. Namely, α , which controls the balance between Wasserstein and Gromov-Wasserstein losses – i.e. how important functional data is compared to anatomical data – is set to 0.5. Empirically, we see that $\alpha = 0.5$ yields values for the Wasserstein loss which are larger than that of the Gromov-Wasserstein loss, meaning that functional data drives these alignments. Secondly, ρ , which sets the importance of marginal constraints – i.e. to what extent more or less mass can be transported to / from each voxel – is set to 1. Empirically, this value leads to all voxels being transported / matched with equal importance. Finally, ε , which controls for entropic regularization – i.e. how blurry computed alignments will be – is set to 10^{-4} . Empirically, this value yields alignments which are anatomically very sharp, i.e source voxels are matched with a handful of target voxels only (and vice-versa).

6.2.2 Decoding

Brain input There is a time *lag* between the moment a stimulus is played and the moment it elicits a maximal BOLD response in the brain [Glover, 1999]. Moreover, the effect induced by this stimulus might span over multiple consecutive brain volumes. To account for these effects, we use a standard Finite Impulse Response (FIR) approach. It fits the decoder on a time-shifted, multi-volume version of the BOLD response. The time shift we implement on the BOLD response is equivalent to adding a time lag to the stimuli. In particular, we refer to the number of brain volumes to aggregate together in the FIR approach as the *window size*. Different *aggregation functions* can be used, such as stacking or averaging. Figure 11.2 describes these concepts visually.

Video output The matrix of latent features \mathbf{Y} is obtained by using a pre-trained image encoder on each video frame and concatenating all obtained vectors in \mathbf{Y} . Similarly to Ozelik and VanRullen, 2023, and as illustrated in Figure 6.1.A, we seek to predict CLIP 257×768 (high-level) and VD-VAE (low-level) latent representations. We use visual – as opposed to textual – CLIP representations [Radford et al., 2021]. For comparison, we reproduce our approach on latent representations from CLIP CLS (high-level) and AutoKL (low-level), which happen to be much smaller³ and are computationally easier to fit.

Model Fitting the decoder consists in deriving $\mathbf{W} \in \mathbb{R}^{v,m}$, $\mathbf{b} \in \mathbb{R}^m$ the solution of a Ridge regression problem – i.e. a linear regression with L2 regularization – predicting \mathbf{Y} from \mathbf{X} .

Evaluation We evaluate the performance of the decoder with retrieval metrics. Let us denote $\mathbf{X}_{\text{train}}$ and $\mathbf{Y}_{\text{train}}$ the brain and latent features used to train the decoder, \mathbf{X}_{test} and \mathbf{Y}_{test} those to test the decoder, and $\hat{\mathbf{Y}} \triangleq \mathbf{W}\mathbf{X}_{\text{test}} + \mathbf{b}$ the predicted latents. We ensure that the train and test data are disjoint.

We randomly draw a retrieval set K of 499 frames without replacement from the test data. In other words, negative frames come from the stimuli that were shown to the participant for fMRI sessions used at test time. For each pair $(\hat{\mathbf{y}}, \mathbf{y})$ of predicted and ground truth latents, one derives their cosine similarity score $s(\hat{\mathbf{y}}, \mathbf{y})$, as well as similarity scores to all latents \mathbf{y}_{neg} of the retrieval set $s(\hat{\mathbf{y}}, \mathbf{y}_{\text{neg}})$. Let us denote $r_K(\hat{\mathbf{y}}, \mathbf{y})$ the rank of \mathbf{y} , which we define as the number of elements of K whose similarity score to $\hat{\mathbf{y}}$ is larger than $s(\hat{\mathbf{y}}, \mathbf{y})$. In order for the rank to not depend on the size of K , we define the *relative rank* as $r(\hat{\mathbf{y}}, \mathbf{y})/|K|$. Finally, one derives the median relative rank $\text{MR}(\hat{\mathbf{Y}}, K)$:

$$r_K(\hat{\mathbf{y}}, \mathbf{y}) \triangleq |\{\mathbf{y}_{\text{neg}} \in K \mid s(\hat{\mathbf{y}}, \mathbf{y}_{\text{neg}}) > s(\hat{\mathbf{y}}, \mathbf{y})\}|$$

$$\text{MR}(\hat{\mathbf{Y}}, K) \triangleq \text{median}\left(\left\{r_K(\hat{\mathbf{y}}, \mathbf{y})/|K|, \forall (\hat{\mathbf{y}}, \mathbf{y})\right\}\right)$$

³Dimensions for CLIP CLS: 768 ; CLIP 257×768 : $257 \times 768 = 197\,376$; AutoKL: $4 \times 32 \times 32 = 4\,096$; VD-VAE: $2 \times 2^4 + 4 \times 2^8 + 8 \times 2^{10} + 16 \times 2^{12} + 2^{14} = 91\,168$

6.2.3 Decoding and alignment setups

Within- vs out-of-subject Let us consider a decoder trained on data $(\mathbf{X}_{\text{train}}^{S_1}, \mathbf{Y}_{\text{train}}^{S_1})$ from a given participant. The *within-subject* setup consists in testing it on left-out data $(\mathbf{X}_{\text{test}}^{S_1}, \mathbf{Y}_{\text{test}}^{S_1})$ acquired in the same participant. The *out-of-subject* setup consists in testing it on data $(\mathbf{X}_{\text{test}}^{S_2}, \mathbf{Y}_{\text{test}}^{S_2})$ acquired in a left-out participant. Note that in this setup, test stimuli can be different from training stimuli, or not.

Single- vs multi-subject The *single-subject* setup consists in training a decoder using data from one participant only. The *multi-subject* setup consists in training a decoder using data from multiple participants. In this study, data from several participants are stacked, resulting in a matrix $\mathbf{X}_{\text{multi}} \in \mathbb{R}^{n_1 + \dots + n_p, v}$ and $\mathbf{Y}_{\text{multi}} \in \mathbb{R}^{n_1 + \dots + n_p, m}$, where p is the number of participants.

Aligned vs un-aligned Let S_1 be the *reference* participant. In the out-of-subject and multi-subject setups, data coming from different participants can be *functionally aligned* – or not – to that of the *reference* participant. It modifies these respective setups as follows: (1) in the out-of-subject case, it corresponds to aligning S_2 onto S_1 , such that a decoder trained on S_1 will be tested on $\mathbf{X}_{\text{test}}^{S_2 \rightarrow S_1}$, $\mathbf{Y}_{\text{test}}^{S_2}$, (2) in the multi-subject case, all participants are aligned to S_1 and the decoder is trained on a concatenation of $\mathbf{X}^{S_1}, \mathbf{X}^{S_2 \rightarrow S_1}, \dots, \mathbf{X}^{S_p \rightarrow S_1}$ (see notations introduced at the end of section 6.2.1) and $\mathbf{Y}^{S_1}, \dots, \mathbf{Y}^{S_p}$.

Setups of interest are visually described in Figure 6.3.A.

Evaluation under different data regimes Note that the alignment and decoding models do not need to be fitted using the same amount of data. In particular, we are interested in evaluating out-of-subject performance in setups where a lot of data is available for the reference participant, and little data is available for the left-out participant: this would typically be the case in clinical setups where, usually, little data is available in patients. In this case, we evaluate whether it is possible to use this small amount of data to align the left-out participant onto the reference participant, and have the left-out participant benefit from a decoder previously trained on a lot of data.

6.2.4 Datasets

We analyze two fMRI datasets. The first dataset [Wen et al., 2017] comprises 3 human participants who watched 688 minutes of video. The videos consists of 18 train segments of 8 minutes each and 5 test segments of 8 minutes each. Each training segment was presented twice. Each test segment was presented 10 times. Each segment consists of a sequence of roughly 10-second video clips. The fMRI data was acquired at 3T, 3.5mm isotropic spatial resolution and 2-second temporal resolution. It was minimally pre-processed with the same pre-processing pipeline as that of the Human Connectome Project [Glasser et

al., 2013]. In particular, data from each participant are projected onto a common volumetric anatomical template. Similarly to prior work on this dataset [Kupersmidt et al., 2022; Wang et al., 2022; Wen et al., 2018], we use runs related to the first 18 video segments - 288 minutes - as training data, and runs related to the last 5 video segments as test data.

The second dataset [Allen et al., 2022] – denoted as the Natural Scenes Dataset (NSD) – comprises 8 participants who are shown 10 000 static images three times. They were scanned over 40 sessions of 60 minutes, amounting to 2 400 minutes of data. Instead of raw BOLD signal, we leverage precomputed per-trial regression coefficients accessible online. See supplementary section 11.6 for more details.

6.2.5 Preprocessing

For the Wen et al., 2017 dataset, we implement minimal additional preprocessing steps for each participant separately. For this, we (1) project all volumetric data onto the FreeSurfer average surface template *fsaverage5* [Fischl, 2012], then (2) regress out cosine drifts in each vertex and each run and finally (3) center and scale each vertex time-course in each run. Figure 11.1 gives a visual explanation as to why the last two steps are needed. The first two steps are implemented with Nilearn [Abraham et al., 2014]⁴ and the last one with Scikit-Learn [Pedregosa et al., 2011].

Additionally, for a given participant, we try out two different setups: a first one where runs showing the same video are averaged, and a second one where they are stacked.

The Allen et al., 2022 dataset is already preprocessed by the original authors, and maps of beta coefficients from a General Linear Model are accessible online.

6.2.6 Hyper-parameters selection for decoders

To train decoders, we use the same regularization coefficient α_{ridge} across latent types and choose it by running a cross-validated grid search on folds of the training data. We find that results are robust to using different values and therefore set $\alpha_{\text{ridge}} = 50\,000$. Similarly, values for lag, window size and aggregation function are determined through a cross-validated grid search.

	D	S1	S2	S3	
CLIP 257 × 768	50.0	9.4	6.8	7.8	MR ↓
	1.0	13.8	16.4	13.6	Acc ↑
VD-VAE	50.0	29.9	30.2	28.5	MR ↓
	1.0	3.0	3.5	3.1	Acc ↑
CLIP CLS	50.0	15.1	10.6	11.0	MR ↓
	1.0	8.4	10.5	9.9	Acc ↑
AutoKL	50.0	24.9	21.8	26.0	MR ↓
	1.0	3.9	3.8	3.3	Acc ↑

Table 6.1: Within-subject metrics for all participants and all latent types on the test set Reported metrics are relative median rank ↓ (MR) of retrieval on a set of 500 samples, top-5 accuracy % ↑ (Acc) of retrieval on a set of 500 samples. Chance level is at 50.0 and 1.0 for these two metrics respectively. These results were averaged across 50 retrieval sets, hence results are reported with a standard error of the mean (SEM) smaller than 0.01. The *Dummy* (D) model systematically predicts the mean latent representation of the training set and achieves chance level.

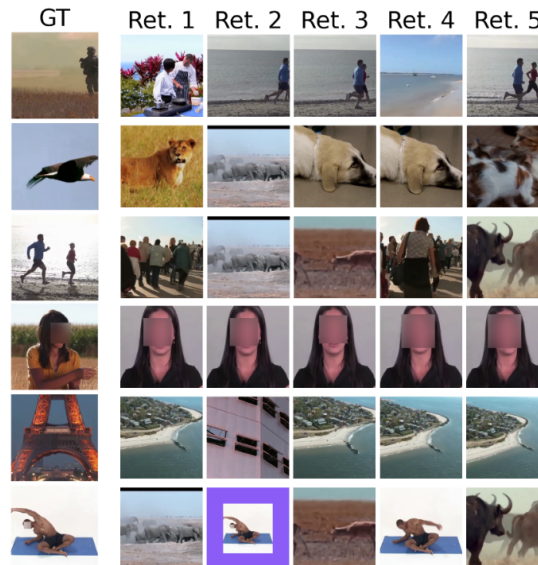


Figure 6.2: Image retrievals using predicted latent representations of CLIP 257 × 768 latents

We use a model fitted on Subject 2 (S2) and predict the latent representation of unseen videos (test set). Ground truth (GT) images featured within the first 5 retrieved (Ret.) images are indicated with a bold purple border. In a given row, images which appear similar across columns are actually different frames of the same video clip. Images featuring human faces were blurred. More cases are available in supplementary Figure 11.4.

6.3 Results

6.3.1 Within-subject prediction of visual representations from BOLD signal and retrieval of visual inputs

We report video decoding results on the retrieval task in Table 6.1. For all three participants of the Wen 2017 dataset, and for all four types of latent representa-

⁴<https://nilearn.github.io>

tions considered, a Ridge regression fitted within-subject achieves significantly above-chance performance. Besides, performance varies across participants, although well-performing participants reach good performance on all types of latents.

Results reported in Table 6.1 were obtained for a lag of 2 brain volumes (i.e. 4 seconds since $TR = 2$ seconds) and a window size of 2 brain volumes that were averaged together (see definitions in section 6.2.2). These parameters were chosen after running a k-fold cross-validated grid search for lag values ranging from 1 to 5, a window size ranging from 1 to 3, and 2 possible aggregation functions for brain volumes belonging to the same window (namely averaging and stacking). Figure 11.3 shows results using the averaging aggregation function for different values of lag and window size, averaged across participants. These results were obtained by stacking all runs of the training dataset, as opposed to averaging repetitions of the same video clip. The two approaches yielded very similar metrics. We give more details in section 6.3.3.

Figure 6.2 shows retrieved images for Subject 2. Qualitatively, we observe that retrieved images often fit the theme of images shown to participants (with categories like indoor sports, human faces, animals, etc.), yet with occasional failures.

6.3.2 Out-of-subject decoding and multi-subject training

As illustrated in Figure 6.3, models trained on one participant do not generalise well to other participants: using CLIP 257×768 , the within-subject and out-of-subject median rank (MR) are respectively 8.0 and 17.2 on average. However, functional alignment allows to reduce the median rank back to 11.1 on average. In particular, we show that left-out participants do not need to have the same amount of available data as training participants to benefit from their decoder: with only 30 minutes of data, left-out participants can reach performance which would have required roughly 100 minutes of data in a within-subject setting.

In addition, we show that a single decoder trained on all functionally aligned participants can reach better results than a decoder trained on all un-aligned participants (MR is 7.7 against 8.6 averaged across subjects), and performs on par with each corresponding single-subject decoders.

Framework generality Note that, in Figure 6.3, we chose the best performing participant (S2) as the reference participant. We report all other combinations of reference and left-out participants in Supplementary Figures 11.6 and 11.7 and find that all effects persist for all combinations. In addition, supplementary Figures 11.8 and 11.9 show that these results hold for all types of latents. Furthermore, we replicate this experiment on four participants from the Natural Scenes Dataset: decoding performance, reported in Tables 11.1 and 11.2, is higher than for the Wen 2017 dataset, probably due to using much more training data. Importantly, functional alignment allows the median rank to drop from

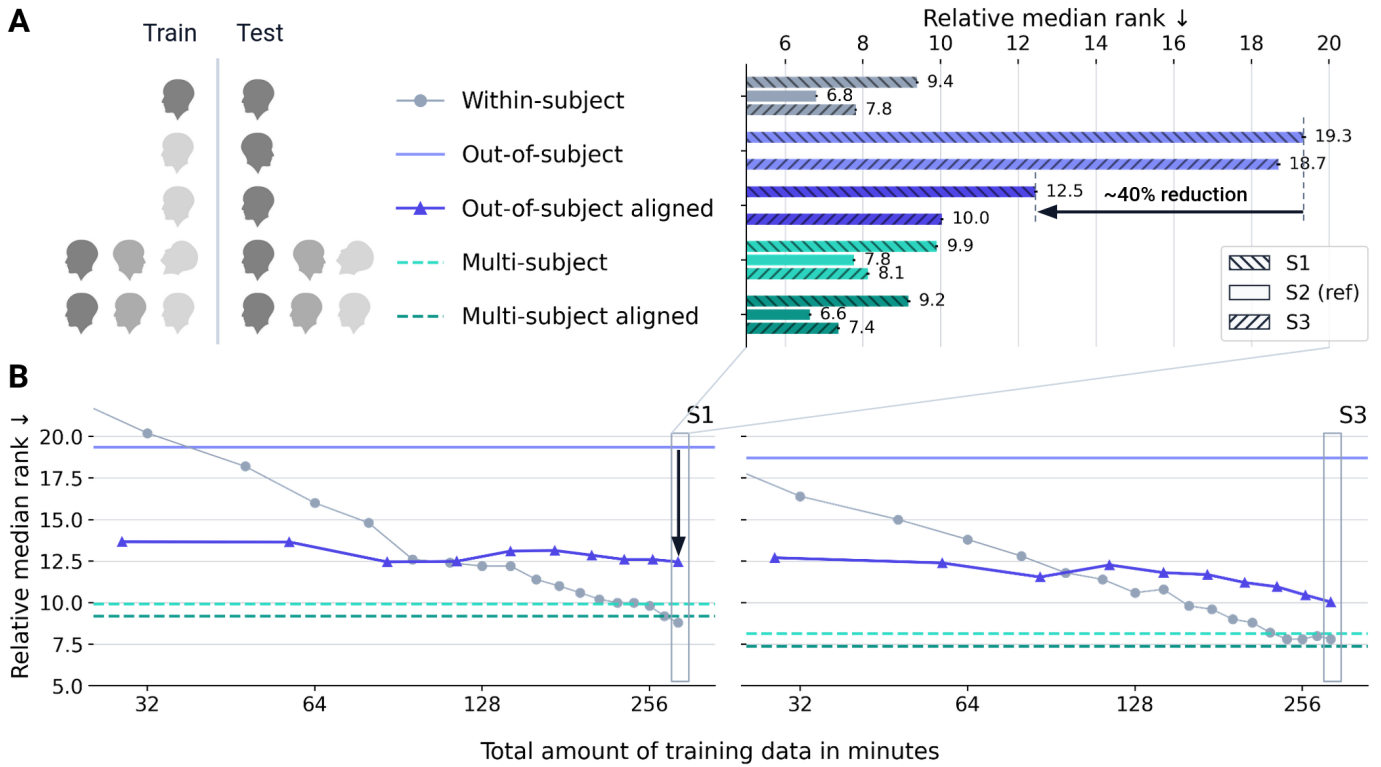


Figure 6.3: Effects of functional alignment on multi-subject and out-of-subject setups

We report relative median rank ↓ in all setups described in section 6.2.3 for CLIP 257×768 . In all *aligned* cases, S1 and S3 were aligned onto S2. In all *out-of-subject* cases, we test S1 and S3 onto a decoder trained on S2. In all *multi-subject* cases, the decoder was trained on all data from all 3 participants. **A.** In this panel, all models (alignment and decoding) were trained on all available training data. Results for other latent types are available in Figure 11.8. **B.** In left-out S1 and S3, decoding performance is much better when using functional alignment to S2 (solid dark purple) than when using anatomical alignment only (solid pale purple). Performance increases slightly as the amount of data used to align participants grows, but does not always reach levels which can be achieved with a single-subject model fitted in left-out participants (solid pale gray dots) when a lot of training data is available. Training a model on multiple participants yields good performance in all 3 participants (dashed pale teal) which can be further improved by using functional alignment (dashed dark teal). Results for other latent types are available in Figure 11.9.

22.5 (baseline) to 11.5 on average in the out-of-subject setup for CLIP 257×768 . In particular, it drops to 8.3 on average across left-out participants when S7 is used as reference. Finally, Figures 11.11 and 11.12 present all other setups we ran for this study. In particular, they show that a multi-subject aligned model (e.g. trained on S1 and S2) has better performance on aligned left-out participants (e.g. S3) than a single-subject model (e.g. trained on S2 only).

Exploring computed inter-subject alignments To better understand how brain features are transformed by functional alignment, we show in Figure 6.4 how vertices from participant S1 are warped to fit those of participant S2. To this end, we colorise vertices in S1 using the MMP 1.0 atlas [Glasser et al., 2016] and use $\phi_{S1 \rightarrow S2}$ to transport each of the three RGB channels of this coloring to S2. Note that both participants' data lie on fsaverage5.

We see that, in low data regimes, FUGW does not recover a smooth inter-subject mapping of the cortical surface, but still manages to recover the cortical organization of the occipital lobe. A greater amount of data allows FUGW to reconstruct inter-subject mappings that are anatomically consistent in a much higher number of cortical areas such as the temporal and parietal lobes, and, unexpectedly, in the primary motor cortex as well. The prefrontal cortex and temporo-parietal junction (TPJ) seem challenging to map, perhaps due to greater inter-subject functional variability or lesser responsivity in those regions.

6.3.3 Influence of training set size and test set repetitions

Recent publications [Ozcelik and VanRullen, 2023; Scotti et al., 2023; Tang et al., 2023] in brain decoding using fMRI have shown impressive results, but these results are obtained using unusually large datasets and signal-to-noise ratios (e.g. tens of hours of 7T fMRI per participant). To evaluate the importance of these two factors, we report in Figure 6.5 performance metrics for models trained with various amounts of data and tested with various amounts of noise.

Firstly, using a fixed test set, our results allow to systematically estimate the quantity of training data needed to achieve a given decoding performance. Interestingly, our results show that stacking two runs displaying the same stimuli yields better results than averaging them. Besides, for a given acquisition budget, showing different stimuli (as opposed to repeating stimuli) yields small but systematic performance improvements.

Secondly, reported performance metrics only hold in favorable signal-to-noise setups. Indeed, the test set associated with the Wen 2017 dataset comes with 10 runs for each video segment, which, when averaged together, greatly reduce the noise level. However, as reported in Figure 6.5, when tested in real-life signal-to-noise conditions (i.e. only one run per test video clip), our models' performance degrades: when using CLIP latents, for each participant, it is approximately twice as bad as when averaging all 10 runs.

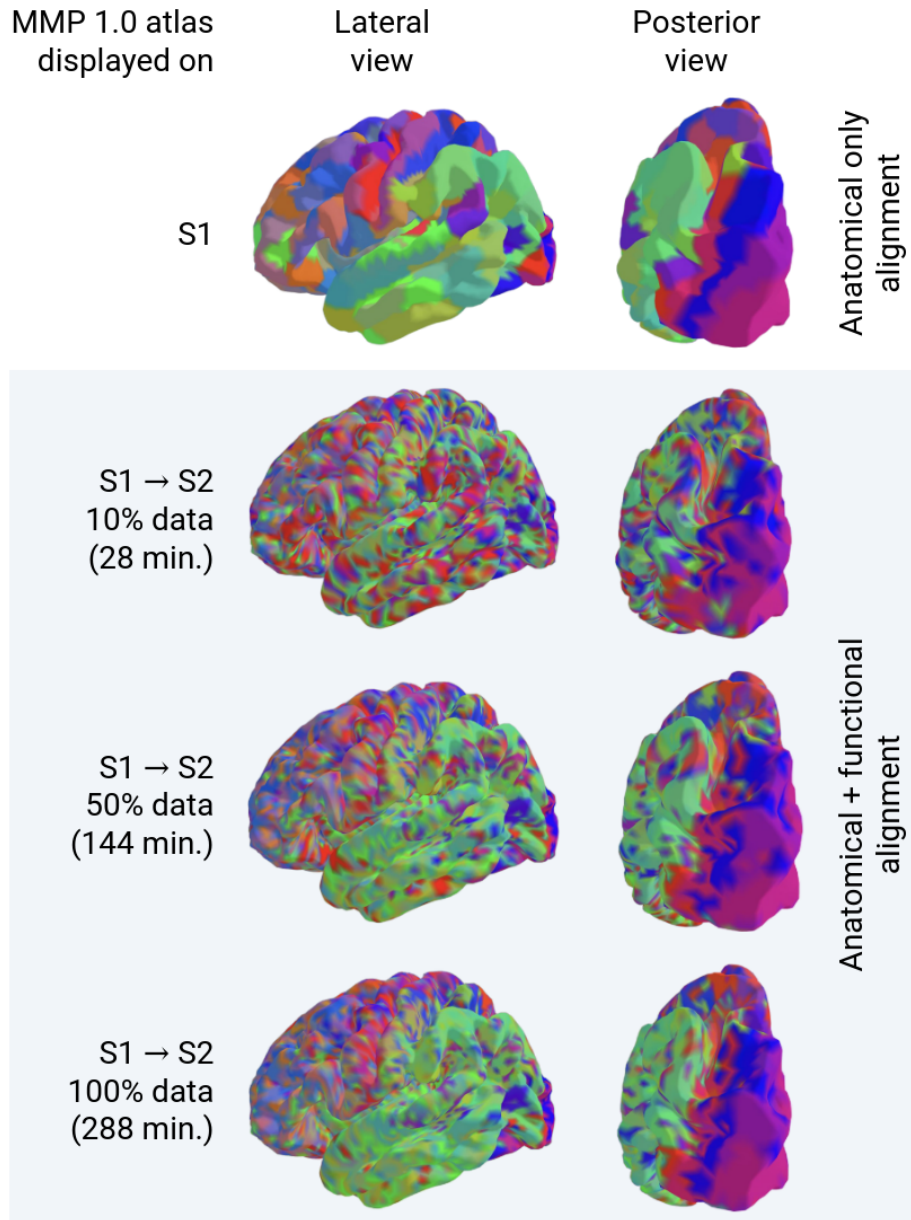


Figure 6.4: Visualizing functional alignments in the left hemisphere Vertices of the left-out participant (top row) are warped by FUGW. The result of this transport is visualized on the reference participant (rows 2, 3, and 4). Fitting FUGW with increasing amounts of data gradually leads the inter-subject mapping to better respect the cortical organisation of multiple areas, including non-visual ones. Note that all 3 models were fitted using the same number of iterations.

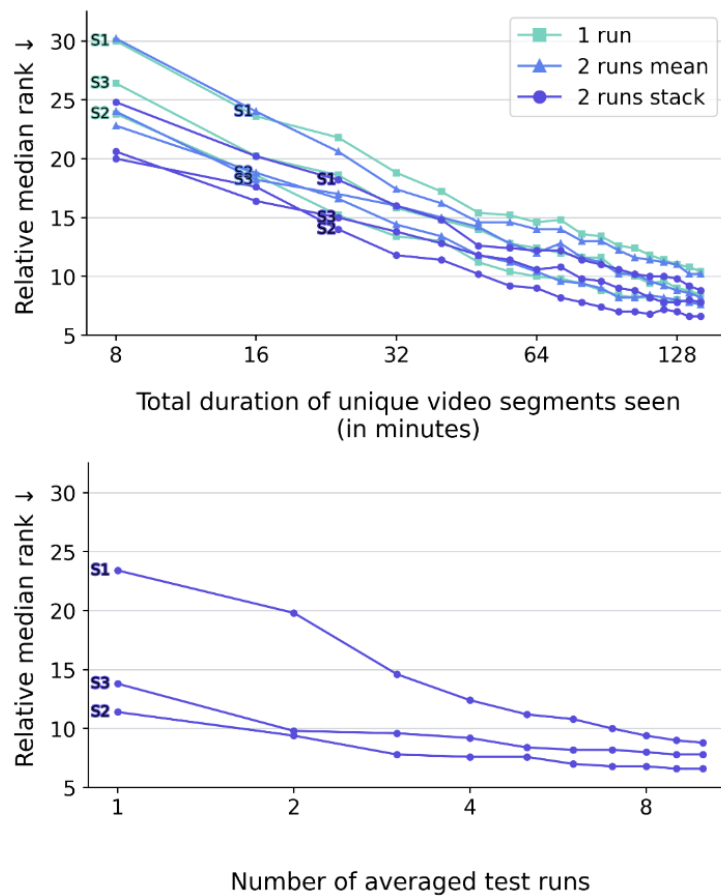


Figure 6.5: Influence of training set size and test set noise Relative median rank \downarrow on a fixed test set gets better as more training data is used to fit the model (top). Interestingly, averaging brain volumes of 2 similar runs does not bring improvements compared to using just 1 run. Instead, stacking runs yields significant improvements. Note that training sets using 2 runs have twice as much data as those using 1 run. Finally, these metrics are highly affected by the noise level of the test set (bottom): averaging more runs in the test set yields better metrics despite using the same decoder.

6.4 Discussion

Impact The present work confirms the feasibility of using fMRI signals in response to natural images and videos to decode high level visual features [Nishimoto et al., 2011]. It further demonstrates that it is possible to leverage these fMRI signals to estimate meaningful functional alignments between participants, and use them to transfer semantic decoders to novel participants.

In particular, our study shows that decoding brain data from a left-out participant can be substantially improved by aligning this left-out participant to a large reference dataset on which a decoder was trained. Our method thus paves the way to using models trained on large amounts of individual data to decode signals acquired in smaller neuroimaging studies, which typically record an hour or two of fMRI data for each participant [Madan, 2022].

In addition, this study reports decoding accuracy in setups where participants are shown test stimuli for the first time, thus providing insight into how these models would perform in real-time decoding. While performance improves with the number of repetitions at test time, reasonable decoding performance of semantics can be achieved with only one repetition in two out of three participants.

Lastly, by systematically quantifying decoding accuracy as a function of the amount of training data, the present work brings insightful recommendations as to what stimuli should be played in future fMRI datasets collecting large amounts of data in a limited number of participants. In the current setup (naturalistic movie clips acquired at 3T), training with diverse semantic content is more valuable than training with repeated content for fitting decoding models.

Limitations This work is a first step towards training accurate semantic decoders which generalize across individuals, but subsequent work remains necessary to ensure the generality of our findings.

Firstly, although the reported gains in out-of-subject setups are significant, the small number of participants present in the dataset under study requires replications on larger cohorts. However, to our knowledge, no other dataset has presented similar features to [Wen et al., 2017](#), namely a large amount of data per participant and a large variety of video stimuli.

Secondly, our approach currently requires left-out participants to watch the same stimuli as reference participants. It is yet unclear whether functional alignment could bring improvements without this constraint. However, multi-subject decoding can probably help partially address this issue: since it is possible to train a decoder on multiple participants and because not all of them have to watch the same movies, it is possible that a lot of different movies - each seen by a different participant used in the training set - could be used as “anchors” for left-out individuals.

Finally, while restricting this study to linear models makes sense to establish baselines and ensure replicability, non-linear models have proved to perform as well [[Scotti et al., 2023](#)], and constitute a natural improvement of this work.

Ethical implications Out-of-subject generalization is an important test for decoding models, but it raises legitimate concerns. In this regard, this study highlights that signal-to-noise ratio still currently makes it challenging to very accurately decode semantics in a real-time setup, and that a non-trivial amount of data is needed per individual for these models to work. In particular, it would be interesting to see if recent work in perception decoding in MEG [[Benchetrit et al., 2023](#); [Défossez et al., 2023](#)] could be applied to out-of-subject setups with a method similar to ours. Moreover, we stress that, while great progress has been made in decoding perceived stimuli, imagined stimuli are still very challenging to decode [[Horikawa and Kamitani, 2017](#)]. Nonetheless, it is important for advances in this domain to be publicly documented. We thus advocate that

open and peer-reviewed research is the best way forward to safely explore the implications of inter-individual modeling, and more generally brain decoding.

Chapter 7

An application of FUGW to whole-brain inter-species functional alignment

In this chapter, we leverage all contributions from previous chapters to compute functional alignment transport plans between human and non-human primates, and test their relevance through a transfer learning task. Namely, we train a brain decoder on human participants and test it on macaque data aligned to human data. Figure 7.1 illustrates the general pipeline of this study.

We show that the brain decoder can successfully predict the latent features of images seen by the macaque and that the predicted latents can be used to generate captions describing the images seen by the macaque.

7.1 Methods

The methods used for this study are very similar to that of Chapter 6. The key differences are that (1) the brain decoders now consist of lightweight deep-learning models instead of linear regressions and (2) the lag for macaque data is two seconds – instead of four for humans – to account for the differences in hemodynamic and MION response functions.

Nonetheless, in the following paragraphs, we provide a detailed description of the methods used in this study.

7.1.1 MRI data

Stimuli Two types of stimuli are used in this study. First, the *Monkey Kingdom* movie is shown to all subjects. It is an eighty-minute documentary showing macaques interacting freely while a narrator describes their actions in English.

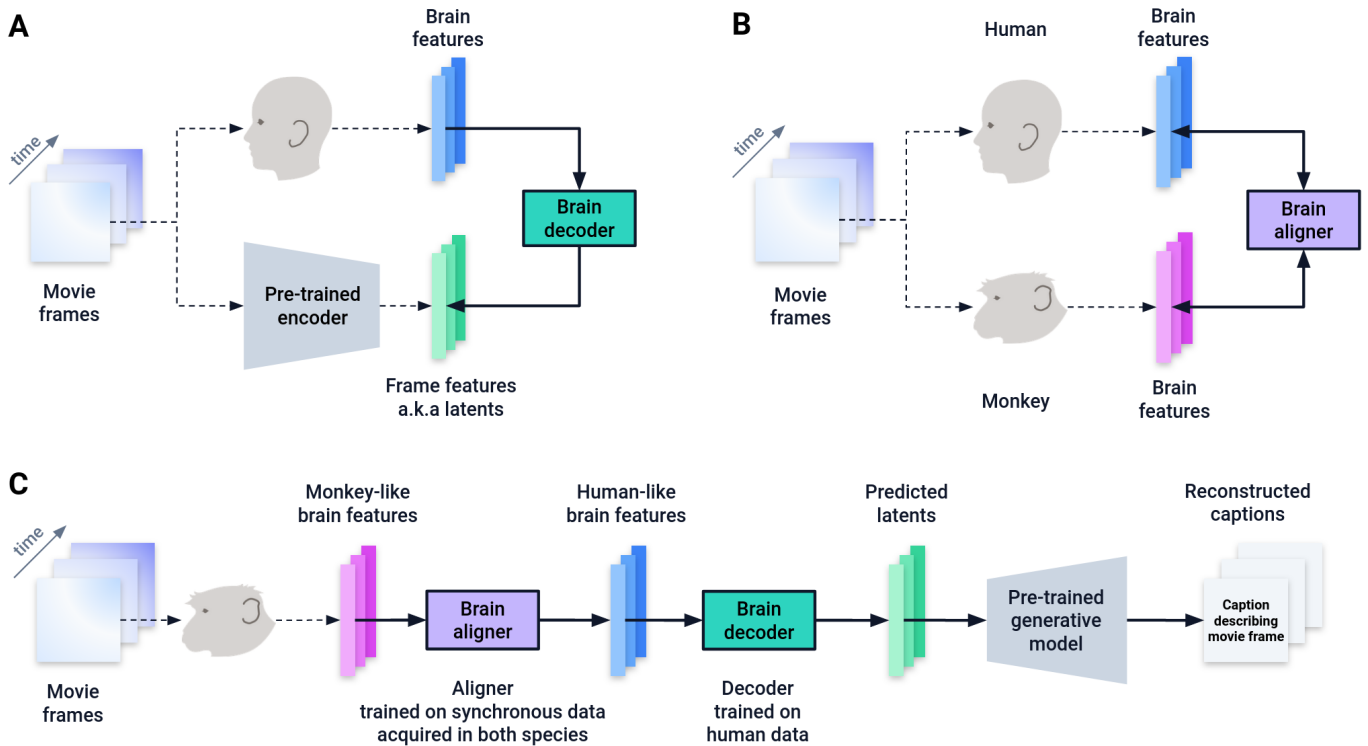


Figure 7.1: Inter-species semantics decoding **A.** fMRI responses, denoted as brain features, are recorded while a subject is watching movies or movie-clips. A semantically-rich representation of seen frames is obtained using a pre-trained image encoder. These frame features are denoted as latents. A brain decoder is trained to predict these latents from brain responses. **B.** fMRI responses collected in different individuals watching the same movies can be used to align their cortical activation patterns. **C.** By successively using a brain aligner and a brain decoder, it is possible to decode brain-activity across species. Moreover, predicted latents can be used as inputs of other pre-trained generative models. In particular, text-generative models can be used to generate short captions describing the movie frames seen by the monkey.

The movie was divided into five consecutive sixteen-minute parts shown during different fMRI runs. Secondly, movie clips from [Nishimoto et al., 2011](#) were shown to human participants. These clips consist of short (10-20 seconds) excerpts from movies, presented in a random order, and organised in runs of eleven minutes. Runs are divided into two groups: 12 training runs – denoted as Clips-Train – containing different movie clips without repetitions, and 9 validation runs – denoted as Clips-Valid – whose movie clips are similar to the training runs, repeated about 3 to 4 times.

Monkey data Two macaques (1 female) were injected with MION [[Leite et al., 2002](#)] and scanned at 3T while performing a fixation task. Over a few days, they were shown the five runs of Monkey Kingdom multiple times (4 to 5) in a random order.

Human data We use a subset of individuals from the Individual Brain Charting dataset [[Pinho et al., 2018](#)]. It comprises MRI acquisitions for a large variety of stimuli - including movie watching - in a limited number of participants. We sub-select all 8 humans of the dataset who have seen all previously mentioned stimuli. They were scanned at 3T while performing a fixation task for the movie clips stimuli, and gazing freely for Monkey Kingdom. The five runs of Monkey Kingdom were shown only once.

Data preprocessing Regarding humans, we refer the interested reader to the original paper of IBC [[Pinho et al., 2018](#)] for a detailed description of the data preprocessing. Regarding macaques, the fMRI images were slice-time corrected, motion corrected and co-registered to an anatomical template (MNI for humans, MEBRAINS for macaques). Individual meshes were created for each subject using the Freesurfer pipeline, and individual fMRI data were projected on these meshes.

7.1.2 Brain decoder

Input brain features Let us denote as $\mathbf{X} \in \mathbb{R}^{T,V}$ the matrix of brain features used as inputs, where T is the number of acquired brain volumes, and V is the number of brain vertices. We strive to minimally process the brain features to keep the model as simple as possible. Namely, for each run of each subject in both species, we (1) regressed out cosine drifts and (2) standardized each vertex's time series to have zero mean and unit variance. Additionally, in macaques, we averaged runs showing the same movie part to increase the signal-to-noise ratio. Besides, to predict movie frames seen when brain volume t was acquired, we aggregate p consecutive brain volumes starting at brain volume $t + l$. We refer to $p, l \in \mathbb{N}$ as the *window size* and *lag* respectively. Based on the work presented in Chapter 6 [[Thual et al., 2023](#)], when using signal acquired in human participants, we aggregate brain volumes by taking their mean, set a window

size of two brain volumes and a lag of four seconds. We used the same hyper-parameters for the macaques, except for the lag, which was set to two seconds to account for the differences in hemodynamic and MION response functions. Moreover, when using signal acquired in macaques as input of a brain decoder trained in human participants, we use the opposite of the signal to account for the fact that the peak of the HRF in humans is inverted compared to the MION response function.

Output latent features Let us denote as $Y \in \mathbb{R}^{T,M}$ the matrix of latent features to be predicted by the model, where T the number of brain volumes and M is the dimension of the latent space. For a given brain volume x_t , the associated latent vector y_t is the average of $(y_{t,k})_{1 \leq k \leq K}$, the respective latent representations of K evenly-spaced frames seen when the brain volume was acquired. Each $y_{t,k}$ is the output of the frame through a pre-trained image encoder. We rely on the ViT-L/14 architecture and checkpoints distributed with the original CLIP paper [Radford et al., 2021]. This visual encoder was trained on a large set of images from the web and encodes high-level features of images. Note that, in this study, *CLIP* refers to the encoding model used to generate the ground truth latents, while *Clips-Train/Valid* refers to a set of movie clips shown to human participants.

Model Similarly to previous work [Scotti et al., 2023], we train a deep learning model to predict latent features from brain features. However, we use a much lighter architecture for simplicity and because our training dataset is smaller. Namely, our architecture consists of a few residual blocks constituting the *backbone*. The output of the backbone is then fed into a *retrieval head* that minimises a contrastive loss (MixCo). An architecture more similar to that of Scotti et al., 2023 would also feed the backbone's output to a *retrieval head* that minimises a reconstruction loss (Mean Squared Error for instance), which we do not attempt here. Our contrastive head is a simple projector made of multiple linear layers with GELU activations. We strive to keep the number of parameters low to avoid overfitting. In particular, we use a backbone with 2 residual blocks, each containing 512 hidden units. The head consists of 2 linear layers with 512 hidden units.

Augmentations Contrastive models are known to benefit from data augmentations [T. Chen et al., 2020]. To augment the predicted latent representations, we compute augmentations on the ground truth images and use these augmented images as inputs of the encoder. We use the following augmentations: random crop and resize, random colour jiggle, and random gaussian blur. We pre-compute 20 augmentations, from which we randomly sample during training.

Training The model is trained using the AdamW optimiser with a learning rate of 10^{-4} , a batch size of 128, a weight decay of 0, and a dropout rate of 0.8. We train the model for 20 epochs and keep the model with the best validation loss.

Evaluation We evaluate the performance of the decoder on a retrieval task. First, let $\mathbf{X}_{\text{test}}, \mathbf{Y}_{\text{test}}$ be the brain features and latent features of held-out runs, respectively. We define a set K of n random samples drawn from \mathbf{Y}_{test} without replacement. Then, we predict the latent features using our pre-trained model on test brain features $\hat{\mathbf{Y}} \triangleq f(\mathbf{X}_{\text{test}})$. Each predicted sample $\hat{\mathbf{y}} \in \hat{\mathbf{Y}}$ is compared to each sample $\mathbf{y}_{\text{neg}} \in K$ using a similarity function s – we use the cosine similarity. Eventually, as defined in Equation 7.1, we also compute the similarity between the predicted and ground truth latent $(\hat{\mathbf{y}}, \mathbf{y})$, and the rank of \mathbf{y} is defined as the number of samples in K whose similarity is greater. Finally, we report the median of the relative ranks across all predicted samples.

$$\begin{aligned} r_K(\hat{\mathbf{y}}, \mathbf{y}) &\triangleq |\{\mathbf{y}_{\text{neg}} \in K \mid s(\hat{\mathbf{y}}, \mathbf{y}_{\text{neg}}) > s(\hat{\mathbf{y}}, \mathbf{y})\}| \\ \text{MR}(\hat{\mathbf{Y}}, K) &\triangleq \text{median}\left(\{r_K(\hat{\mathbf{y}}, \mathbf{y})/|K|, \forall (\hat{\mathbf{y}}, \mathbf{y})\}\right) \end{aligned} \quad (7.1)$$

Hyper-parameters tuning Given the large number of hyper-parameters, we did not perform an exhaustive search. Instead, we tuned hyper-parameters sequentially, each time keeping the best values found so far. We used the validation runs of the movie clips (Clips-Valid) of the human participants to tune the hyper-parameters on decoders fitted on the training runs of the movie clips (Clips-Train).

Text generation Predicted latents can be used as inputs for other pre-trained generative models. In particular, text-generative models can be used to generate short captions describing the movie frames seen by the monkey. We rely on ClipCAP [Mokady et al., 2021], a lightweight model based on GPT-2 that generates captions from image latent representations obtained with CLIP ViT models. We train from scratch the model on CLIP ViT/14 latent representations from images of the COCO dataset by relying on the implementation of ClipCAP available on Github¹.

Similarly to Ozelik and VanRullen, 2023, we standardise predicted latents to have zero mean and unit variance, before scaling them to the mean and variance of the latents seen during training, and finally use them as inputs for the text-generative model.

7.1.3 Brain alignment

Input brain features Given left-out and a reference subject who might be from different species, we seek to derive a mapping $\mathbf{P} \in \mathbb{R}^{N,P}$ between the N and P vertices of the left-out and reference brains, respectively. To this end, we rely

¹https://github.com/rmokady/CLIP_prefix_caption

on the work from [Thual et al., 2022, 2023]. This work uses optimal transport to compute a soft permutation of the vertices of the left-out individual that maximises the similarity between functional features of the reference individual, while preserving the anatomical organisation of the cortex.

Formally, let $\mathbf{X}^{\text{out}} \in \mathbb{R}^{T,N}$ be the matrix of brain functional features of the left-out subject, where T is the number of acquired brain volumes, and N is the number of brain vertices. Let $\mathbf{D}^{\text{out}} \in \mathbb{R}^{N,N}$ be the matrix of geodesic distances between vertices of the left-out subject. Finally, let $\mathbf{w}^{\text{out}} \in \mathbb{R}^N$ be the vector of vertex-wise weights, which are used to balance the contribution of each vertex to the alignment. Reciprocally, we define $\mathbf{X}^{\text{ref}} \in \mathbb{R}^{T,P}$, $\mathbf{D}^{\text{ref}} \in \mathbb{R}^{P,P}$, and $\mathbf{w}^{\text{ref}} \in \mathbb{R}^P$ for the reference subject.

\mathbf{P} is such that it minimises the loss function $\mathcal{L}_\Theta(\mathbf{P})$ described in Equation 7.2:

$$\mathcal{L}_\Theta(\mathbf{P}) \triangleq (1 - \alpha) \mathcal{L}_W(\mathbf{P}) + \alpha \mathcal{L}_{\text{GW}}(\mathbf{P}) + \rho \mathcal{L}_U(\mathbf{P}) + \varepsilon H(\mathbf{P}) \quad (7.2)$$

Gromov Wasserstein loss
Wasserstein loss
Marginal constraints
Regularization

where:

- $\mathcal{L}_W(\mathbf{P}) \triangleq \sum_{\substack{0 \leq i < N \\ 0 \leq j < P}} \|\mathbf{X}_i^{\text{out}} - \mathbf{X}_j^{\text{ref}}\|_2^2 P_{i,j}$
- $\mathcal{L}_{\text{GW}}(\mathbf{P}) \triangleq \sum_{\substack{0 \leq i,k < N \\ 0 \leq j,l < P}} |D_{i,k}^{\text{out}} - D_{j,l}^{\text{ref}}|^2 P_{i,j} P_{k,l}$
- $\mathcal{L}_U(\mathbf{P}) \triangleq \Phi(\mathbf{P}_{\#1} \otimes \mathbf{P}_{\#1} | \mathbf{w}^{\text{out}} \otimes \mathbf{w}^{\text{out}}) + \Phi(\mathbf{P}_{\#2} \otimes \mathbf{P}_{\#2} | \mathbf{w}^{\text{ref}} \otimes \mathbf{w}^{\text{ref}})$
- $H(\mathbf{P}) \triangleq \Phi(\mathbf{P} \otimes \mathbf{P} | (\mathbf{w}^{\text{out}} \otimes \mathbf{w}^{\text{ref}}) \otimes (\mathbf{w}^{\text{out}} \otimes \mathbf{w}^{\text{ref}}))$

Here, $\Phi(\cdot | \cdot)$ denotes a divergence – either the Kullback-Leibler divergence when aligning humans to humans, or the L2-norm when aligning macaques to humans – $\mathbf{P}_{\#1} \triangleq (\sum_j P_{i,j})_{0 \leq i < n}$ is the first marginal of \mathbf{P} , $\mathbf{P}_{\#2} \triangleq (\sum_i P_{i,j})_{0 \leq j < n}$ is the second marginal of \mathbf{P} , $\alpha \in [0, 1]$, $\rho \in \mathbb{R}_+$ are the hyper-parameters setting the relative importance of each constraint, and $\Theta \triangleq (\mathbf{X}^{\text{out}}, \mathbf{X}^{\text{ref}}, \mathbf{D}^{\text{out}}, \mathbf{D}^{\text{ref}}, \alpha, \rho, \varepsilon)$.

Hyper-parameters selection We compute human-to-human alignments using the same hyper-parameters as in Thual et al., 2022, 2023, namely $\alpha = 0.5$, $\rho = 1$, $\varepsilon = 10^{-4}$, and Φ is the Kullback-Leibler divergence. We compute macaque-to-human alignments setting Φ to the L2-norm and study the impact of α , ρ , and ε on the performance of the brain decoder.

Feature projector Subsequently, we define $\phi: \mathbf{X} \mapsto (\mathbf{P}^T \mathbf{X}^T) \oslash \mathbf{P}_{\#2} \in \mathbb{R}^{N,P}$ where \oslash is the element-wise division, a function that transports any matrix of brain features from the left-out participant to the reference participant.

7.2 Results

7.2.1 Training a semantic decoder on multiple human participants

Although the stimuli used to train the decoders are not as rich as that of other decoding studies – the present work features about two hours per participant, when [Scotti et al., 2023](#) were using more than 30 hours, [Thual et al., 2023](#) more than 5 – we successfully trained a brain decoder on multiple human participants from the IBC dataset.

We present the retrieval performance of our brain decoder on one human participant in [Figure 7.2](#). The decoder was trained on all IBC subjects aligned onto sub-04 using the Clips-Train data (both for functional alignment and fitting the brain decoder) and tested on the Clips-Valid data of the same participant. We fitted the brain decoder hyper-parameters using this setup.

Predictions for random samples – i.e., not selected amongst the best predictions – of the test set can be found in [supplementary Figure 12.1](#).

Results on the validation set show that, for our retrieval task, the latent features predicted by our brain decoder are more similar to that of the image currently seen by the participant than they are to other images. Moreover, retrieved images show semantic consistency, with pictures of landscapes, animals, or water being retrieved in groups.

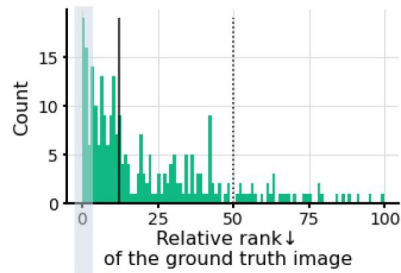
For this particular decoder, the median relative retrieval rank of the ground truth image is 13 when testing on a human participant – chance level is 50. We run a chi-squared test on the distribution of ranks of ground-truth images, which would be uniform at chance level. The test is very significant ($\chi^2 = 5 \cdot 10^2$, p-value $< 10^{-64}$).

7.2.2 Brain decoders with functional alignment transfer from humans to macaques

We freeze the set of hyper-parameters found in the previous section and train a series of brain decoders on more data. Namely, we train eight brain decoders. Each decoder is trained on all IBC subjects. All participants are aligned onto a reference participant using the Clips-Train, Clips-Valid, and the first two segments of Monkey Kingdom data. The same stimuli (Clips-Train, Clips-Valid, and the first two segments of Monkey Kingdom data) are used to train the decoder.

We then test each decoder on the fourth and fifth segments of Monkey Kingdom in left-out aligned macaques and humans. Macaques were aligned to the reference human subject using the first two segments of Monkey Kingdom. We leave the third segment out to ensure no continuity between the third and fourth segments. [Figure 7.3](#) shows that we can successfully transfer each of the eight brain decoders to each of the two macaques. [Figure 7.4](#) shows examples of the best predictions of the brain decoder for one of the macaques. One can see

A



B

Seen image	Caption generated from ground truth latents	Retrieved images	Caption generated from predicted latents
	A close up of a person's leg with a pair of scissors.		A close up of a claw being used to cut a butterfly.
	A person climbing up a steep hill on a red rock.		A man is trying to pull a stunt stunt on a red car.
	A snowy mountain with a mountain range in the background.		A large, dark, tropical forest with a lot of trees.
	A sign that says "Altered Baha'i" on it.		A large yellow star on a large white plant.
	A white dog standing next to a wall.		A man in a white shirt and black tie.
	A cat and a dog are looking at something.		A man in a white shirt and tie holding a toy.
	A close up of a figurine of a mouse.		A man is bending over to pick up a pair of scissors.
	A dog is looking at a figurine of a person.		A man wearing a white dress holding a small animal.
	A close up of a jellyfish on a surface.		A close up of a bunch of scissors sticking out of a plant.
	Two ducks swimming in the water together.		Two people holding a pair of scissors in their hands.

Figure 7.2: Retrieval performance on the validation set of a brain decoder trained on all IBC subjects (A.) Distribution of the relative rank of the ground truth image for a brain decoder trained on Clips-Train using all IBC subjects aligned onto sub-04, and tested on Clips-Valid in sub-04. The bold vertical line indicates the median rank of our decoder. The dotted line indicates the median rank at chance level. **(B.)** Examples for the best predictions of the brain decoder **A.** (Left) The ground truth image and caption generated using the ground truth latents. (Right) The first five retrieved images – i.e., the five images with the highest similarity to the predicted latents – and the caption generated using the predicted latents.

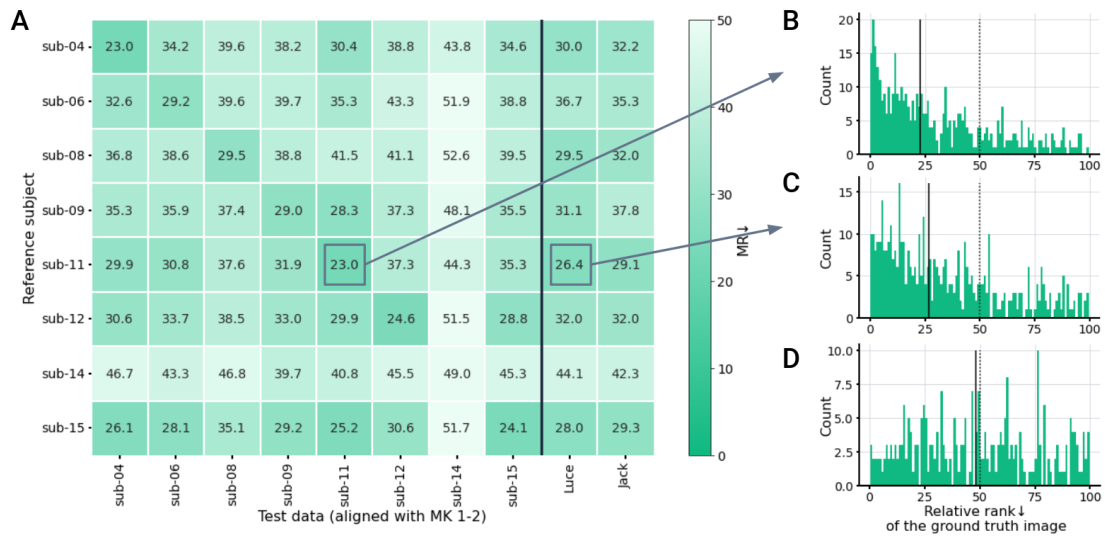


Figure 7.3: Brain decoders trained in human participants transfer to non-human participants **A.** Median relative rank for all 8 participants of IBC and 2 macaques (x-axis) when tested on a brain decoder trained on all IBC participants aligned onto a reference individual (y-axis). **B.** Distribution of the relative rank of the ground truth image in a human participant. **C.** Distribution of the relative rank of the ground truth image in a non-human primate. **D.** Distribution of the relative rank of the ground truth image when predictions for the non-human primate are randomly shuffled.



Figure 7.4: Examples for the best predictions in macaques (Left) The ground truth image and caption generated using the ground truth latents. (Right) The first five retrieved images – i.e., the five images with the highest similarity to the predicted latents – and the caption generated using the predicted latents.

that the best predictions yield latent representations encompassing coherent semantic information because retrieved images are semantically consistent. Predictions for random samples – i.e., not selected amongst the best predictions – of the test set can be found in supplementary Figure 12.2.

For a fair comparison to monkeys, results for human participants were obtained by computing functional alignments to the human reference subject using the first two segments of Monkey Kingdom only. We find that using Clips-Train data to align human data yields better decoding performance, as is reported in supplementary Figure 12.4. It could stem from the fact that, compared to Monkey Kingdom, Clips-Train contains a larger number of samples and that the images shown are more diverse.

Similarly to the previous section, we ran a chi-squared test on each distribution of the ranks of the ground-truth images, which would be uniform at chance level. All tests are very significant, except for sub-14. We also show that the median rank goes up to 50 when the predictions of the brain decoder are randomly shuffled. This control experiment demonstrates that, even across species, the decoder predicts features that correlate with that of the image seen by the participant. Similarly, predicting the average latent features seen during training yields chance level performance.

7.2.3 Evaluating inter-species functional alignments

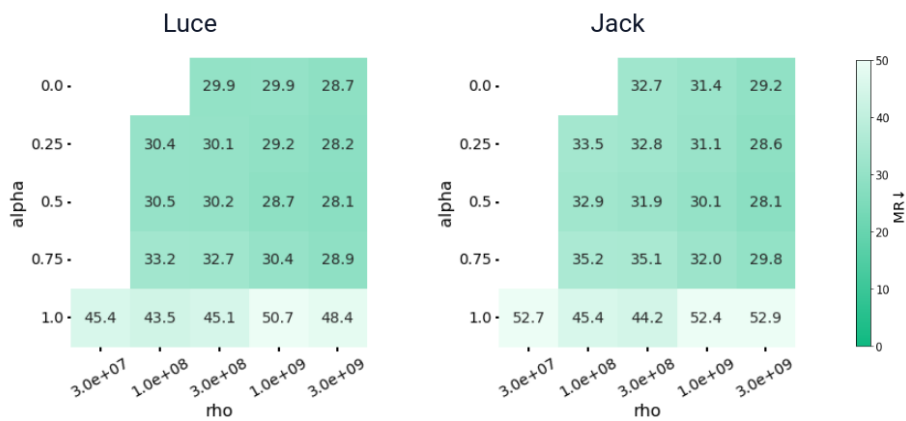
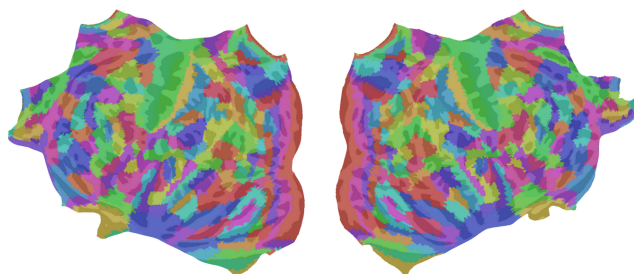


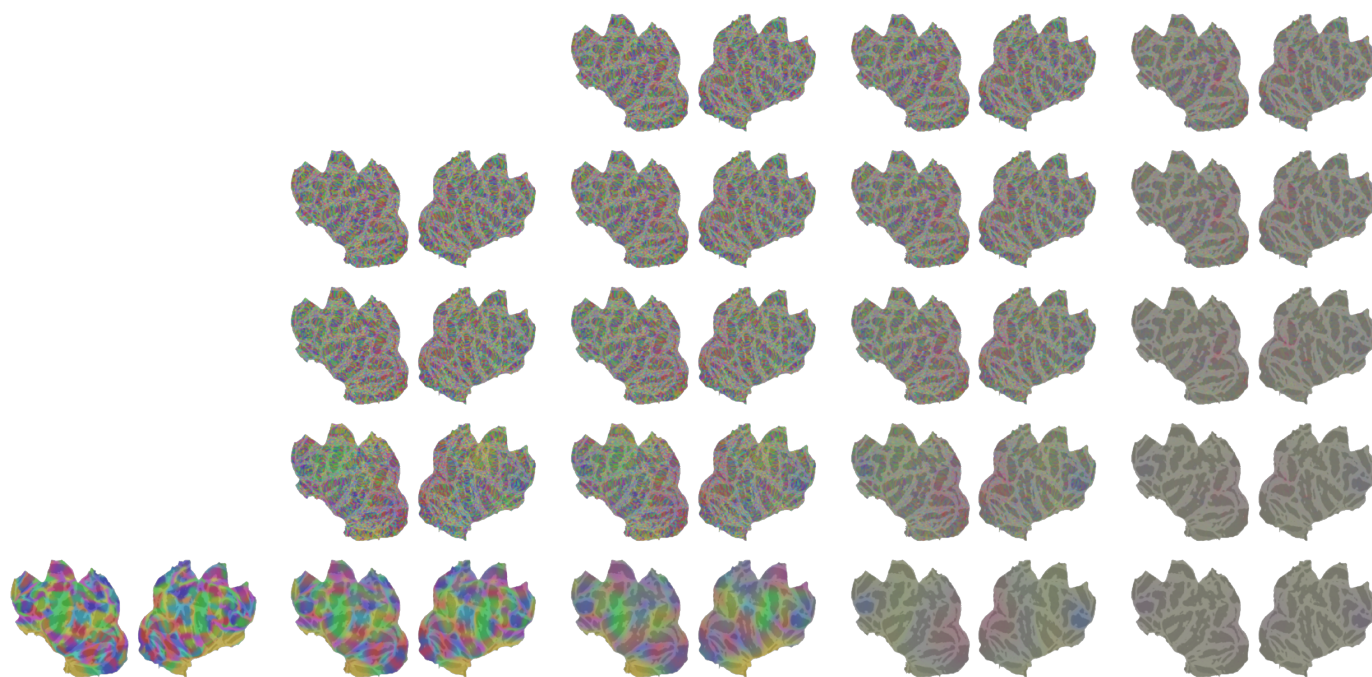
Figure 7.5: Average decoding performance across human-trained decoders for each of the two macaques For each macaque and each pair of (α, ρ) , we aligned the macaque to the human reference subject and evaluated the performance of the brain decoder. We report the average rank across all human reference subjects.

We study the impact of the hyper-parameters α and ρ on the performance of the brain decoder when transferring from humans to macaques and report results in Figure 7.5.

First, we fix $\varepsilon = 100$. We find that for $\alpha = 1$ (i.e., using only anatomical constraints and no functional information), the performance of the brain decoder

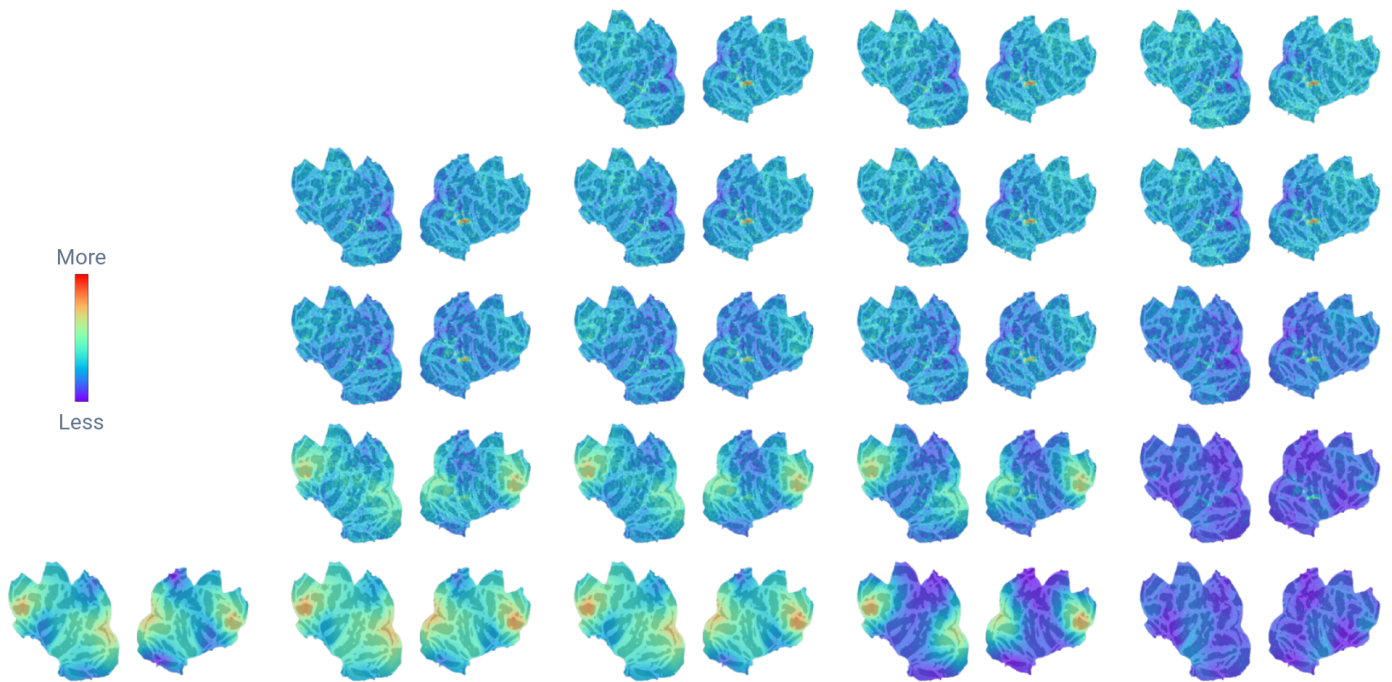


(a) MMP 1.0 atlas [Glasser et al., 2016] in both hemispheres of the human brain displayed on a flattened fsaverage surface.

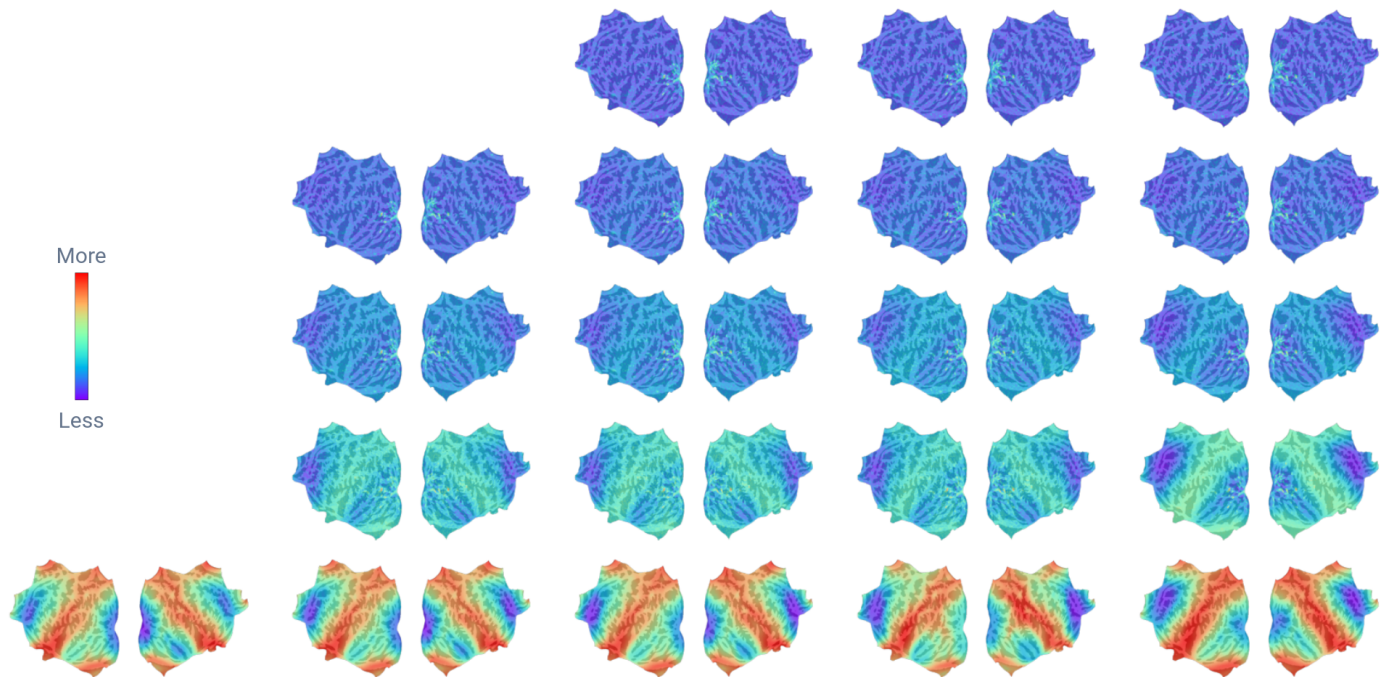


(b) MMP 1.0 atlas transported from a human participant to a macaque using the feature projector ϕ associated with the transport plan \mathcal{P} , for each (α, ρ)

Figure 7.6: Transported atlas for alignments computed between Luce and sub-04 of the IBC dataset For each (α, ρ) reported in Figure 7.5, we display the transported atlas on the macaque brain.

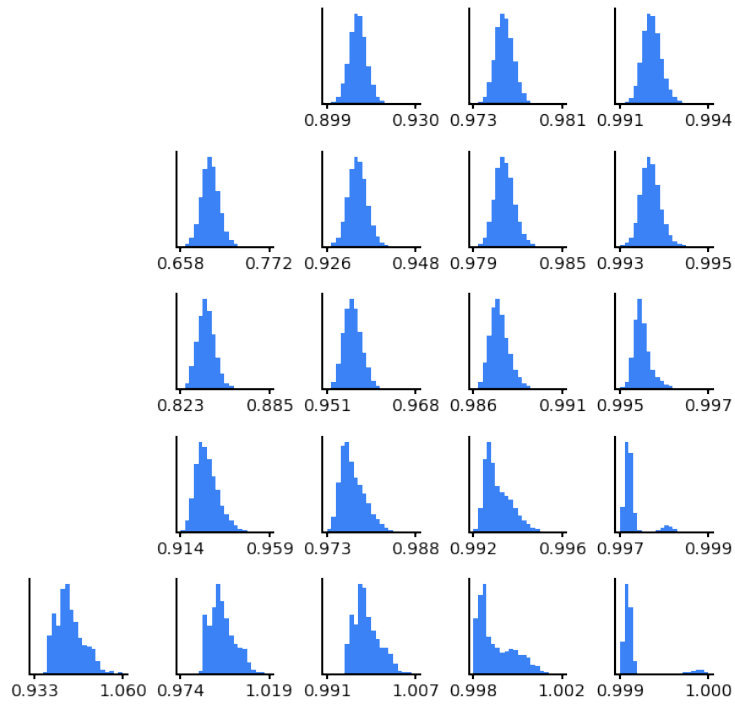


(a) Transported mass for each vertex of the macaque brain for each (α, ρ)

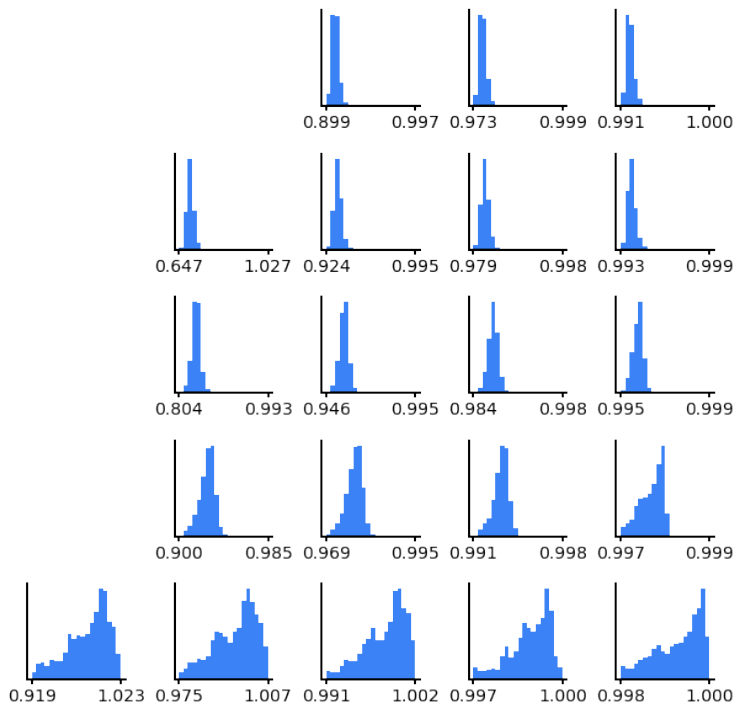


(b) Transported mass for each vertex of the human brain for each (α, ρ)

Figure 7.7: Transported mass in each vertex for alignments computed between Luce and sub-04 of the IBC dataset For each (α, ρ) reported in Figure 7.5, we display the transported mass for each vertex of the macaque and human brains. Scales are reported in Figure 7.8.



(a) Distribution of transported mass for the macaque brain for each (α, ρ)



(b) Distribution of transported mass for the human brain for each (α, ρ)

Figure 7.8: Transported mass distributions for alignments computed between Luce and sub-04 of the IBC dataset We display the distribution of transported mass for the macaque and human brains for each (α, ρ) reported in Figure 7.5.

is at chance level. Whenever $\alpha < 1$, the performance of the brain decoder increases to better-than-chance levels. The best performance is obtained for $\alpha = 0.5$ and $\rho = 3 \cdot 10^9$. Not all combinations of (α, ρ) could be reported, as some combinations led to numerical instabilities during the optimisation of the transport plan: for a given α , the value of ρ needs to be high enough to ensure the convergence of the optimisation algorithm when using the L2-norm as a divergence.

To better understand the transformation induced by the functional alignment, we visualise in Figure 7.6 how it transports a parcellation from a human participant to a macaque. To this end, we use the MMP 1.0 human atlas [Glasser et al., 2016] and transport each RGB channel using the feature projector ϕ defined in the previous section. For each (α, ρ) , we display the transported atlas on the macaque brain in Figure 7.6b.

Empirically, we notice that higher values of ρ lead to smoother transport plans.

Moreover, we visualise in Figure 7.7 the transported mass for each vertex of the human and macaque brains, and the associated mass distributions in Figure 7.8. We find that, in humans, the transported mass is more concentrated in motor areas, and less in visual and prefrontal areas. The opposite is observed in macaques. However, for values of α close to 1, these effects seem to be driven by the Gromow-Wasserstein term – i.e., the anatomical constraint – rather than the Wasserstein term – i.e., the functional constraint – as the same pattern is present for $\alpha = 1$. Values of α close to 0 still exhibit a difference in transported mass in the visual areas, with less mass being transported in the macaque brain and more in the human brain: this could stem from the fact that the visual areas occupy a larger portion of the macaque cortex relative to the total area of the cortex compared to the human cortex.

7.2.4 Choosing the best reference participant for functional alignment

In the previous section, we trained a brain decoder on multiple human participants who were all functionally aligned to a reference participant. First, we find that decoders trained on multiple participants significantly outperform that trained in a single participant. The gain is much larger than that observed in Thual et al., 2023 and could come from using contrastive models instead of linear regressions. However, it is not clear how one should choose the reference participant for functional alignment. We run a series of experiments to investigate this question.

First, we compute brain decoders in a single participant, test them on left-out individuals, and report results in Figure 7.10a. We compute the Pearson correlations between the performance in left-out individuals and (1) the FUGW distance between the left-out and the training participants, and (2) the within-subject performance of the brain decoder (i.e., the performance of the brain decoder trained and tested on the same participant), respectively. In the first case, we

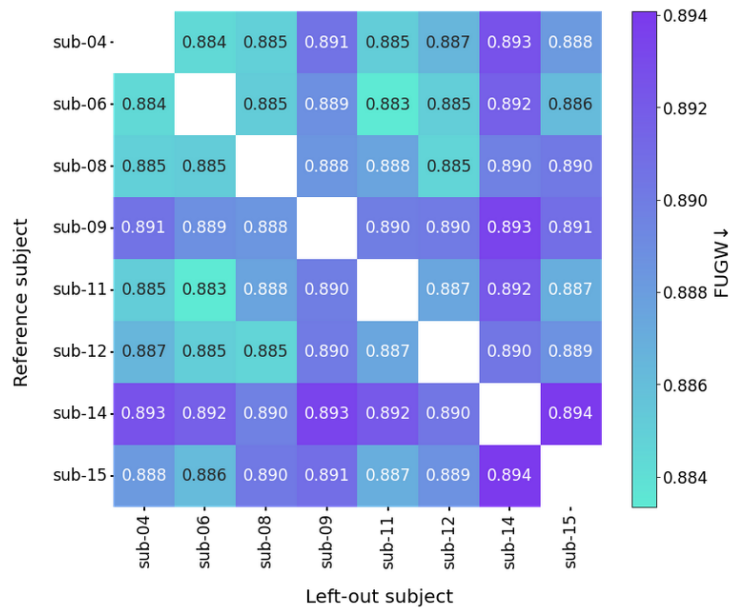


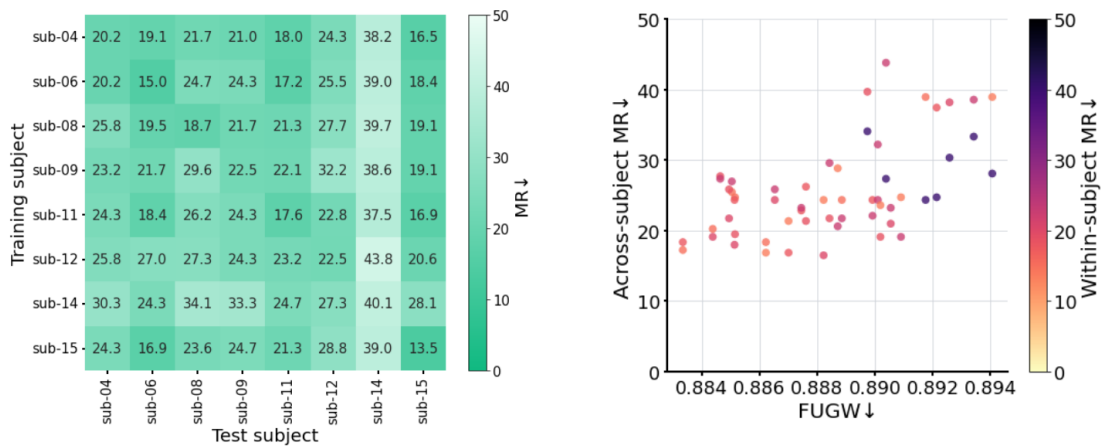
Figure 7.9: FUGW distances between all pairs of IBC subjects on the Clips-Train data For each pair of participants, we fit a transport plan using parameters from [Thual et al., 2022]. We plot the FUGW distance after convergence of the transport plan for each pair of participants. Some individuals have a lower FUGW distance to others in general (see sub-04, sub-06, sub-11), while others have a higher FUGW distance (see sub-09, sub-14).

find a significant positive correlation (correlation = 0.56, p-value = $6 \cdot 10^{-6}$). In the second case, we find a non-significant positive correlation (correlation = 0.2, p-value = 0.1). However, if we remove sub-14 – whose decoding performance is significantly lower than the others for reasons we could not elucidate – both correlations become non-significant.

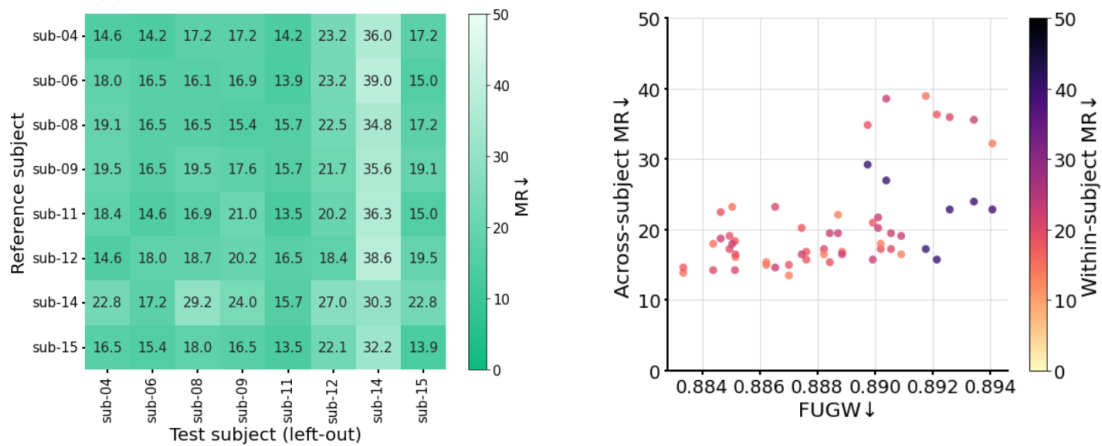
On the contrary, within-subject decoding performance in a given participant very significantly correlates with how well this same subject can be decoded when aligned to other participants (correlation = 0.8, p-value = $5 \cdot 10^{-17}$) even when sub-14 is removed (correlation = 0.6, p-value = $4 \cdot 10^{-6}$). Given the data in this study, we conclude that neither the FUGW distance nor the within-subject performance can predict accurately how a left-out participant can be decoded.

Secondly, we compute brain decoders trained on multiple participants aligned to a reference participant, test them on the left-out individual, and report results in Figure 7.10b. We find the same conclusion as with the previous experiment.

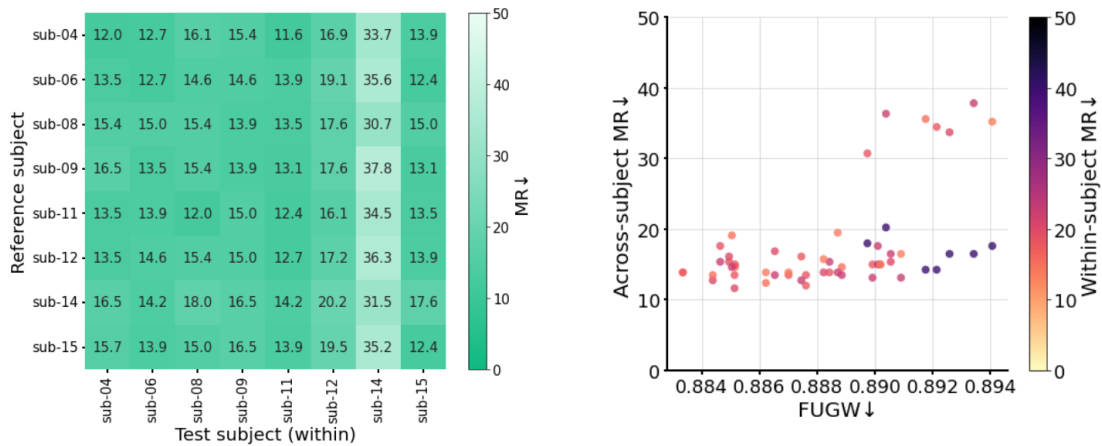
Finally, we compute brain decoders trained on all IBC participants aligned to a reference participant, test them on each training participant, and report results in Figure 7.10c. We find the same conclusion as with the previous experiment.



(a) Decoders are trained on a single IBC participant and tested on each IBC participant.



(b) Decoders are trained on all – except one left-out subject – IBC participants aligned to a single reference participant and tested on the left-out participant.



(c) Decoders are trained on all IBC participants aligned to a single reference participant and tested on each IBC participant.

Figure 7.10: Performance of brain decoders and correlation to FUGW distance and within-subject performance Results for decoders trained in one participant (a), and multiple participants tested in a left-out (b) or within (c) participant. (Left) Median rank. (Right) Median rank as a function of the FUGW distance between the tested participant and the training/reference participant. Colour indicates the within-subject performance of a brain decoder trained on that training/reference participant alone.

7.3 Discussion

This study shows that a decoder of semantic visual perception trained exclusively on human fMRI data can be transferred to non-human primates through functional alignment. These results suggest that the computed inter-species functional alignments contain meaningful information about the similarities and differences in the functional organisation of the cortex between human and non-human primates. Moreover, we advocate that brain decoders should generally be tested in a setup similar to that of this study, i.e., on individuals whose data does not overlap with that of the training set. In our case, the brain decoder was not trained on non-human primate data or the Monkey Kingdom movie segments used for testing.

7.3.1 Limitations

We advocate that the current study is limited by (1) the small number of samples per participant and (2) the nature of the stimuli used. First, we know from other studies [Scotti et al., 2023; Thual et al., 2023] that a greater number of samples per participant should significantly improve the performance of the brain decoder. Moreover, the nature of the stimuli used for training and testing is probably suboptimal. Indeed, brain decoders trained in a single human using Clips-Train seem to exhibit good out-of-subject performance when tested on Clips-Valid on humans who have been functionally aligned using Clips-Train. However, the same brain decoder tested in the same left-out participants, this time using segments of Monkey Kingdom, exhibits a much lower performance. This could stem from the fact that Monkey Kingdom is not as diverse as Clips-Train and Clips-Valid, and that the images shown are not as easy to distinguish. We believe that using richer stimuli could significantly improve the performance of the brain decoder and the relevance of computed inter-species alignments. Lastly, it is not clear whether a fixation task should be imposed on participants.

Moreover, the method presented in this study contains an overwhelming number of hyper-parameters, which can be difficult to tune for users. For the same reason, our current work does not study all combinations of hyper-parameters, which could lead to suboptimal results. However, it proves that at least one combination of hyper-parameters can transfer human-fitted brain decoders to non-human primates using naturalistic stimuli.

7.3.2 Future work

Despite the high number of hyper-parameters, we claim these results can likely be reproduced and improved. Indeed, given the consistency of these results across individuals, we advocate that acquiring a lot of naturalistic stimuli in a very limited number of human and non-human subjects should help derive much more accurate inter-species functional alignments and brain decoders.

Moreover, this work provides baseline hyper-parameters that can be used to fit both the subsequent decoders and alignments.

Chapter 8

Conclusions and opinions

In this last chapter, I wish to adopt a more personal tone to touch upon additional contributions made during this doctorate, discuss the limitations of the methods and results presented in this manuscript, and express personal opinions on the future directions of this research.

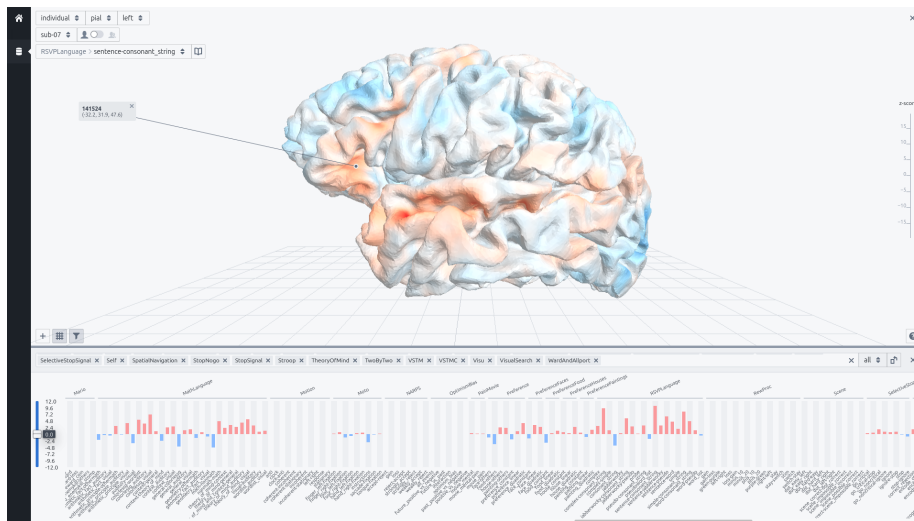
8.1 Additional contributions

8.1.1 Contributions to the analysis of the deep-phenotyping data IBC

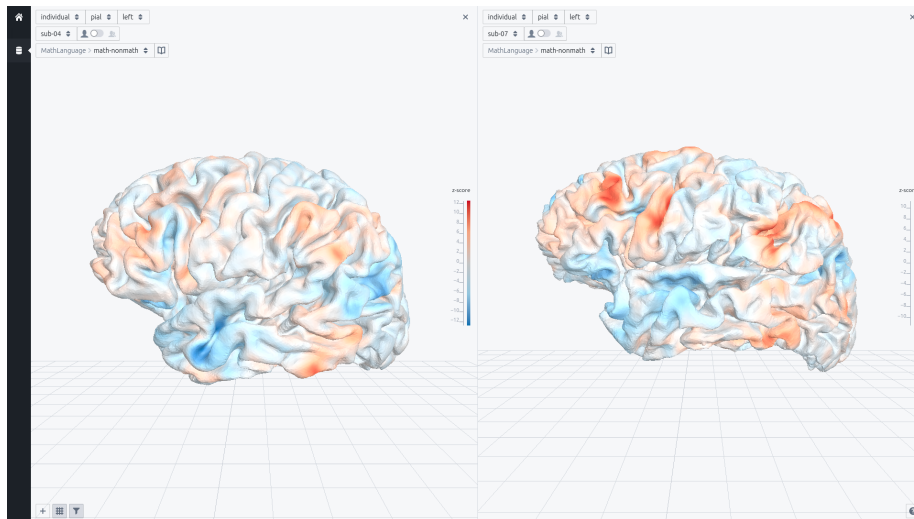
During this doctorate, I was very influenced by the work of the IBC project, which collected a large amount of data from a small number of individuals. I was fortunate to be part of the third release of the dataset [Pinho et al., 2023] and to work from the start of my doctorate with a significant proportion of the collected data that had already been preprocessed. I participated in studies leveraging this data to introduce concepts like that of the *functional fingerprint* of cortical areas [B. Thirion et al., 2021] – namely, the idea that, with enough data in a single participant, one could derive a functional signature that uniquely identifies each part of the cortex – and eventually study how these data can be used to derive individual atlases, and finally how these compare to classical atlases [B. Thirion et al., 2023].

8.1.2 Software contributions to visualisation tools of neuro-imaging data

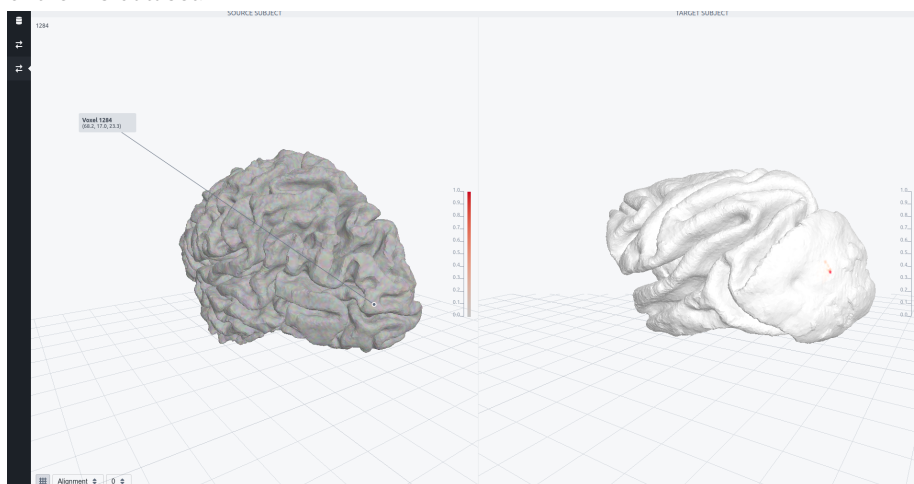
These early studies in my doctorate have led me to develop visualisation tools to better understand the data at hand. Interactively visualising this data was crucial for building intuitions regarding functional variability and functional alignment. I believe that many of the computational errors I have encountered during this doctorate could have been solved quicker by visualising the data under study.



(a) Display how a given vertex in a given IBC participant was significantly activated/deactivated throughout the tasks of the IBC dataset.



(b) Display the same cortical map on the individual cortical surfaces of two participants of the IBC dataset.



(c) Display voxels in the macaque cortex that were matched to a single voxel in the human cortex.

Figure 8.1: Main features of brain-cockpit. The application was designed to ease the following tasks: (a) interactively check the *functional fingerprint* of a given voxel, (b) quickly visualise cortical maps of different participants side by side and (c) interactively explore pre-computed functional alignments.

Brain-cockpit In this context, I first focused on developing¹ and documenting² `brain-cockpit`, an open-source web-based application designed to browse the IBC dataset and `fugw` alignments. It relies on the `three.js` library to render 3D meshes in the browser using WebGL, and on the `react.js` library to manage the state of the application. This setup makes it accessible to any modern web browser without requiring installation. The project also features a `flask` Python back-end that caches and serves pre-computed MRI maps or functional alignments. `brain-cockpit` was designed to ease the following tasks, illustrated in Figure 8.1: (a) interactively check the *functional fingerprint* of a given voxel, (b) quickly visualise cortical maps of different participants side by side and (c) interactively explore pre-computed functional alignments. Moreover, the application makes it possible to send a specific view of the data to a collaborator simply by sharing the current URL.

Human Brain Project deployment `brain-cockpit` also comes with `ansible` scripts to automate deployment on a Linux server. These scripts were used to deploy the application on the Human Brain Project's servers, where it is still accessible to the community and allows users to quickly explore the IBC dataset online³.

Nilearn As a member of the Nilearn core developers team, I had the opportunity to contribute to the development of the library. I mainly focused on enhancing features for the `plotting` module, which is used by a large number of users. My contributions include allowing the plotting of flat cortical maps for fsaverage in any existing resolution, custom cortical background maps – making it possible to plot the sign of the curvature of the cortex, commonly used in other visualisation tools – and custom view angles. I also contributed to the documentation. Finally, I initiated a collaboration with the `niivue`⁴ team to test the feasibility of integrating their javascript 3D renderer in Nilearn, in place of the ones we currently maintain.

8.2 Limitations of the current work

8.2.1 Limitations of the FUGW alignment method

The optimisation problem underlying the FUGW alignment method is well-motivated: functional alignment arises naturally from the Wasserstein loss, anatomical constraints can still be fostered with the Gromov-Wasserstein loss, and marginal constraints enforce a generally well-distributed mass – as was initially showed to be important in Section 2.1. However, FUGW comes with many limitations.

¹<https://github.com/alexisthual/brain-cockpit>

²<https://alexisthual.github.io/brain-cockpit>

³<https://brain-cockpit.tc.humanbrainproject.eu>

⁴<https://github.com/niivue/niivue>

First, the high number of parameters in the FUGW distance makes it unlikely that external users will want to use it if good default values are not provided beforehand. This stresses the importance of running large hyper-parameter searches to find the best values for the parameters α , ρ , and ε in pre-identified setups.

However, hyper-parameters of the FUGW distance are not independent, thus making this search more difficult. Indeed, changing α can also affect the total amount of mass transported, and changing ρ can also affect the relative importance of the entropic regularisation term.

Similarly, the number of vertices/voxels in the source and target spaces can also affect the optimal values of the hyper-parameters.

This difficulty may come from the fact that the Wasserstein and Gromov-Wasserstein terms are inherently not homogeneous: the former is a sum weighted by individual coefficients of the transport plan P , but the latter is a sum weighted by the product of all pairs of coefficients the transport plans P .

8.2.2 Limitations of the data used in the experiments

In this doctorate, we only used the geodesic distance between vertices on the cortical sheet to inform the Gromov-Wasserstein loss. However, white-matter connections between cortical regions may be more informative and shed new light on cortical homologies between human and non-human primates.

More generally, collecting a higher number of samples for functional alignment could yield more insightful conclusions. Moreover, the movie/clips-watching task used in Chapters 6 and 7 may not be as efficient as the static-image viewing task featured in other studies.

8.3 Opinions and future research directions

8.3.1 Advocating for web-based visualisation tools for neuro-imaging data

First, developing open-source software tools is crucial for the scientific community. It has proven to be a powerful way to share knowledge and foster collaborations in the past decades, with some projects gaining major momentum in the neuroscientific community.

Data visualisation, together with data sharing, preprocessing, and analysis, is a critical component of neuroscience. The latter have progressed significantly in the past years, with the development and adoption of shared data standards – such as BIDS – and easy-to-deploy, opinionated processing and analysis tools that come with robust defaults and can be scripted to enable parallelisation.

However, I believe that the former has not seen the same level of progress. Most visualisation tools still require compilation or installation and are challenging

for novice and intermediate developers to contribute to. Moreover, recent massive MRI datasets cannot be downloaded or cached on simple machines, which these pieces of software often require. Web-based solutions constitute a popular and modern alternative to these issues, as they are not as hard to install and distribute to individual machines and allow the use of server-side resources to store and process large amounts of data.

Besides, some have argued that, as a community, we should report whole-brain volumetric unthresholded maps in our publications, in order to limit artefactual effects coming from surface projection and hacking threshold values. I find this idea very interesting, and I believe these maps could come in the form of files stored online and displayed with a generic web-based viewer.

More generally, I think our community needs to switch to server-oriented paradigms for storing, sharing and visualisation purposes, as is the case with the OpenNeuro⁵ and NeuroVault⁶ initiatives. The amount of data can be overwhelming, especially if we are to store fMRI timeseries or raw datasets. I hope that `nii-vue` will become the go-to 3D renderer for neuro-imaging data in the coming years. It would naturally lead to the blooming of web-based visualisation tools with different use cases and levels of complexity, either for use on individual computers or through widely distributed web platforms.

8.3.2 Advocating for larger deep-phenotyping fMRI datasets

Collecting large amounts of data from a limited number of individuals is, in my opinion, a low-risk yet high-gain strategy for the neuro-imaging community. It will not bring new knowledge at the population level but constitutes an easy way to gain invaluable insights into how the brain is organised. Subsequent hypotheses can later be tested at the population level.

I believe these deep-phenotyping datasets should include high-resolution anatomical images of participants, comprising diffusion-weighted images to derive white-matter tracts. Acquiring whole-brain fMRI images, even at the cost of a lower signal-to-noise ratio, is a strategy I would be interested in implementing as it allows prior-free data analyses. Including sub-cortical areas in the acquisition could also prove to be very important.

Besides, ideal functional data should allow one to uniquely identify areas of the cortex without the need for prior anatomical knowledge. Anatomical information would be used afterward to build higher insights into the organisation of the brain across individuals and species.

Finally, I believe that these individuals should be imaged while performing perception tasks over a high number of samples. Moreover, among other criteria, I think that one should choose the type of stimuli based on whether we already have meaningful latent representations for them. For instance, images or gen-

⁵<https://openneuro.org>

⁶<https://neurovault.org>

eral text are stimuli that make it easy to obtain good latent representations using pre-trained decoders. These rich latent representations allow one to rely less on prior knowledge to probe the brain: running a GLM to contrast a specific category of stimuli against others implies defining these categories in advance, which potentially limits the kind of discoveries one can make. Using latent representations can generally alleviate these biases. Moreover, these stimuli can be shown to participants in the form of short samples (i.e., stimuli that last a few seconds at most), and deep-learning models should benefit from this high number of samples. On the contrary, long-context stimuli and mathematical problems, although fascinating, are still challenging to model with existing latent representations.

8.3.3 Advocating for decoding-first approaches

One could argue that the Natural Scenes Dataset [Allen et al., 2022] already constitutes a significant step in giving the community access to a large number of samples collected from a small number of individuals. I strongly support this initiative but also want to express some concern about the way this data has recently been used in widely distributed brain-decoding papers. Indeed, in most of these studies, the data used to test the models consists of beta coefficients derived from a GLM on the entire fMRI session. However, for a given session, some samples acquired during the same runs are used for training and others for testing. In my opinion, it is unclear at this stage how much this artificially boosts the decoding performance of underlying brain decoders.

I think the following questions could help drive the strategy to acquire such datasets: given a brain decoder, how does it perform on data acquired in new fMRI sessions in the same participants? On data sampled from a different distribution than that of the training samples? On data acquired from different individuals? On data acquired with a different fMRI scanner?

In particular, these questions help put in perspective the importance of acquiring data specifically for testing purposes.

More generally, I think it makes sense that our community allocates resources to projects with decoding-first approaches. Indeed, we have rightfully focused on encoding-first approaches in the past decades, which predict brain activity from stimuli properties. However, recent advances in brain-decoding showed that fMRI data might have much more signal than we initially thought. Nonetheless, decoding methods often rely on hard-to-interpret deep-learning models, leading to a tension between encoding approaches that provide grounded results and decoding approaches that seem to indicate there is much more to learn from fMRI data but struggle to draw neuro-scientific conclusions from their results.

However, I want to advocate that decoding-first approaches present real advantages. First, it is an inherently challenging machine-learning task, with much room for improving metrics. As such, it allows for the definition and sharing

of well-posed benchmarks, which I think is trickier to do with encoding approaches. Indeed, predicted measured brain activity highly depends on how the acquisition was performed, as we might not be setting our imaging machines with parameters that are optimal for probing the brain's activity. In other words, the values we are trying to predict do not necessarily contain the information we are looking for. On the contrary, decoding tasks yield better-defined target features, as these are derived directly from the stimuli.

Functional alignment could also be framed as a sub-task of the decoding problem. Indeed, well-performing brain decoders trained on multiple individuals map individual brain activity to a common latent space. With adapted architectures, one could try to invert this mapping using, for instance, normalising flows. It yields a natural way to build individual-to-individual alignments without requiring that these participants see the same stimuli.

Finally, I think it is very unlikely that the current MRI sequences we currently use to acquire data are optimal for decoding purposes. Researching new MRI sequences on their ability to provide features that allow one to train high-performance brain decoders could lead to significant improvements in the field. However, this might require using artificial models of the brain's dynamics to find optimal sequences or, if such models do not exist, to run the subsequent gradient descent in the real world. Cheaper, low-field MRI machines might provide a way to explore this space of hyper-parameters⁷.

⁷<https://www.chipiron.co>

Part III

Appendix

Chapter 9

Supplementary material for Chapter 3 (FUGW)

In the following sections, we provide some additional material to ease the understanding of the underlying alignment problem as well as computational details of solutions of FUGW and FUGW barycenters.

We also show some control experiments during which we used different training data to compute pair-wise alignments and evaluated the proportion of correlation gains that comes from mere signal smoothing.

Eventually, we give details about the IBC dataset (acquisition, preprocessing, fMRI protocols and data splitting).

9.1 Illustration of the alignment problem

We provide in Fig. 9.1 a conceptual illustration of the alignment framework for a pair of subjects.

9.2 Implementation details

MSM configuration We use the default configuration of MSM ¹ and vary parameter *lambda* so as to obtain the best gains in correlation on the test set. We use the same value of *lambda* at each step of MSM and eventually set it to 0.1 after a cross validated grid search.

Correlation gain variability when aligning pairs of subjects Figures 9.2 and 9.3 show correlation gains on the validation and test sets respectively when aligning pairs of subjects from the IBC dataset. Subjects' data was previously

¹MSM default configuration https://github.com/ecr05/MSM_HOCR/blob/master/config/basic_configs/config_standard_MSMPair

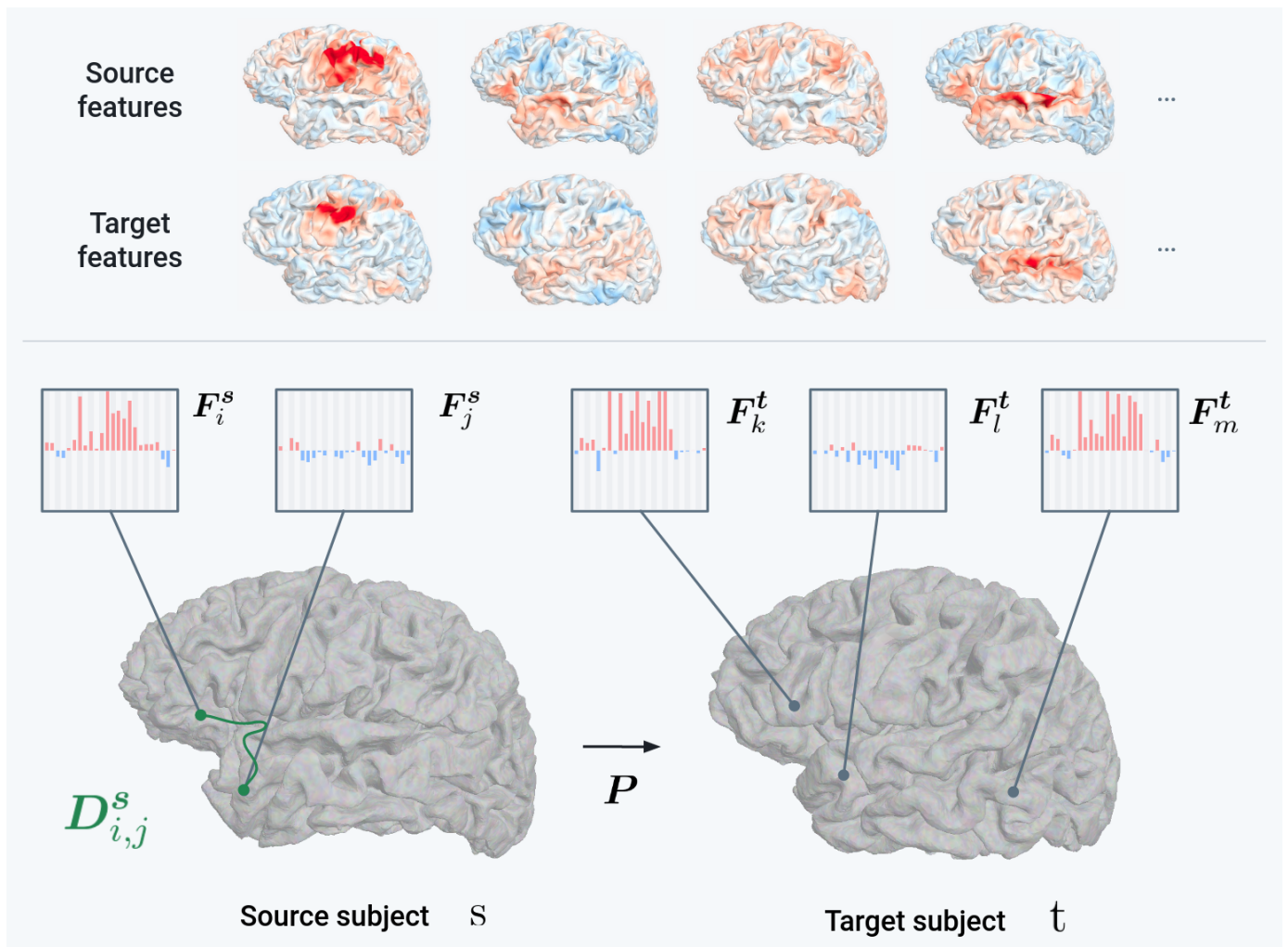


Figure 9.1: Alignment of two brains using functional signatures Using multiple maps of comparable features (left column) for the source and target subjects, we seek to derive an alignment (also referred to as a *coupling*) P that matches parts of the brain with similar features while preserving the global geometry of the cortex.

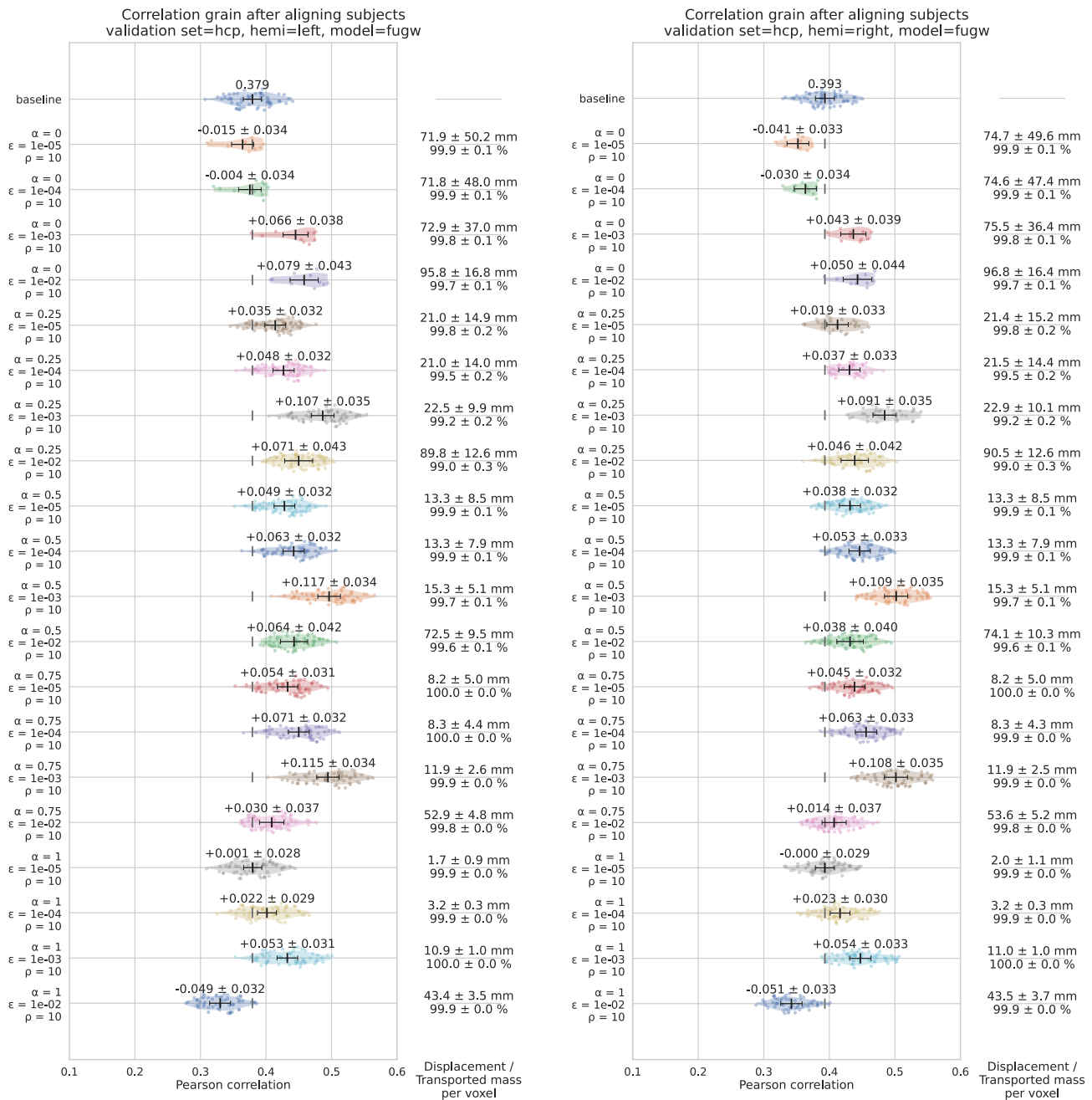


Figure 9.2: Detailed correlation gains on the validation set (HCP tasks), in the balanced case Each line represents a FUGW model trained with different hyperparameters. Each dot represents the mean correlation between contrast maps of the HCP protocol for a given pair of IBC subjects. We compare the average correlation with that of the baseline (top row) where subjects were simply projected on *fsaverage5*. Models for the left hemisphere and right hemisphere are shown respectively on the left and right side.

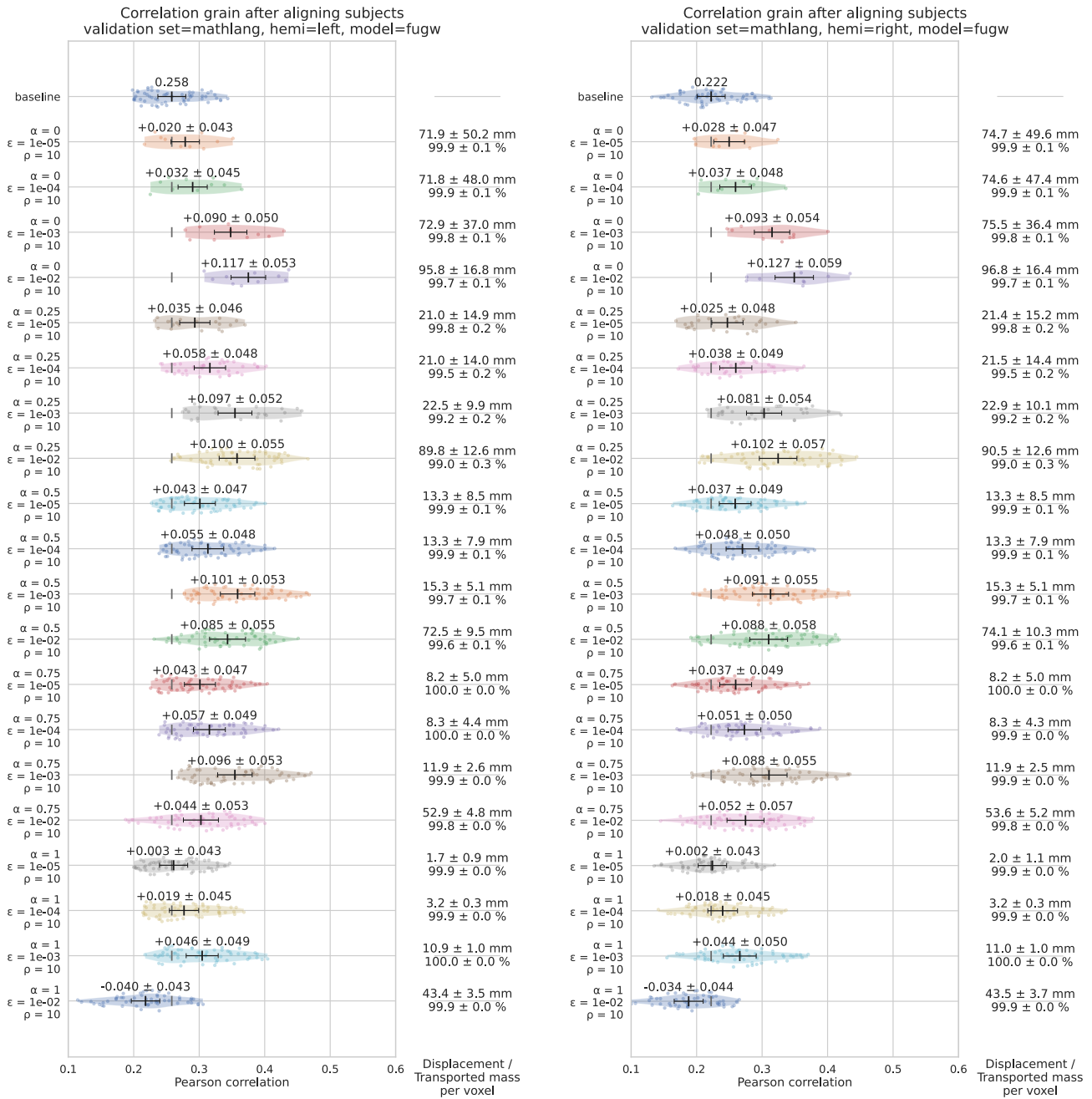


Figure 9.3: Detailed correlation gains on the test set (Mathlang tasks) in the balanced case Similarly to Figure 9.2, each line represents a FUGW model trained with different hyper-parameters. Each dot represents the mean correlation between contrast maps of the Mathlang protocol for a given pair of IBC subjects. We compare the average correlation with that of the baseline (top row) where subjects were simply projected on *fsaverage5*. Models for the left hemisphere and right hemisphere are shown respectively on the left and right side.

projected onto *fsaverage5*. These figures provide us with a better understanding of the standard error and consistency of these gains. Moreover, they show that selection of the best set of hyper-parameters is robust to changing the validation data.

Mesh resolution reduction As mentioned in the core of this paper, aligning meshes with high resolutions can lead to dealing with matrices which won't fit on GPUs. This is typically the case when trying to align two *fsaverage7* hemispheres (160k vertices each) instead of *fsaverage5* hemispheres (10k vertices each).

In order to reduce the number n of aligned vertices, we first group them into small clusters using Ward's algorithm using a method described in [B. Thirion et al., 2014](#). In essence, this method iteratively groups adjacent vertices of a given individual based on feature similarity until I clusters have been formed. Then, for a given cluster u_i of the source subject s , we define its functional signal $\hat{F}_{u_i}^s$ as the mean functional signal of vertices which belong to this cluster. Moreover, for two given clusters u_i and u_j of subject s , we define the anatomical distance \hat{D}_{u_i, u_j}^s between u_i and u_j as the mean geodesic distance between all pairs of vertices between the two clusters (akin to an Energy distance). Eventually, we derive analogous objects \hat{F}^t and \hat{D}^t for the target subject t , and end up in a configuration comparable to that of Experiment 1.

$$\hat{F}_{u_i}^s \triangleq \frac{1}{|u_i|} \sum_{k \in u_i} \mathbf{F}_k^s \in \mathbb{R}^c \quad \hat{D}_{u_i, u_j}^s \triangleq \frac{1}{|u_i| \cdot |u_j|} \sum_{k \in u_i, l \in u_j} D_{k, l}^s$$

Alignment to individual anatomy We qualitatively control that alignments derived between individuals on their individual anatomies make sense in [Figure 9.4](#).

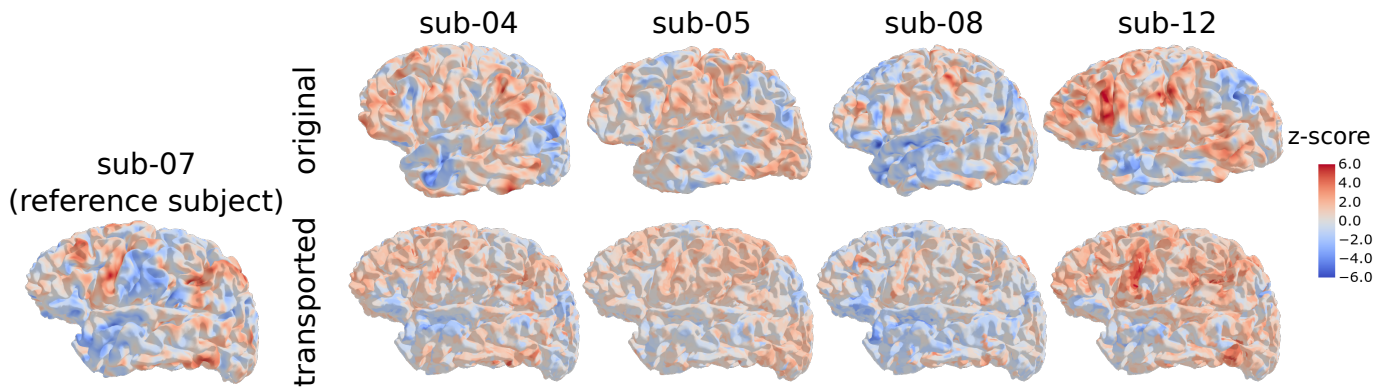


Figure 9.4: Transporting individual maps onto a reference subject FUGW can help bridge the absence of template anatomies and derive pairs of alignments such that all individuals of the cohort are comparable. We display a map taken from the test set contrasting areas activated during mathematical reasoning against areas activated for other stimuli of the protocol.

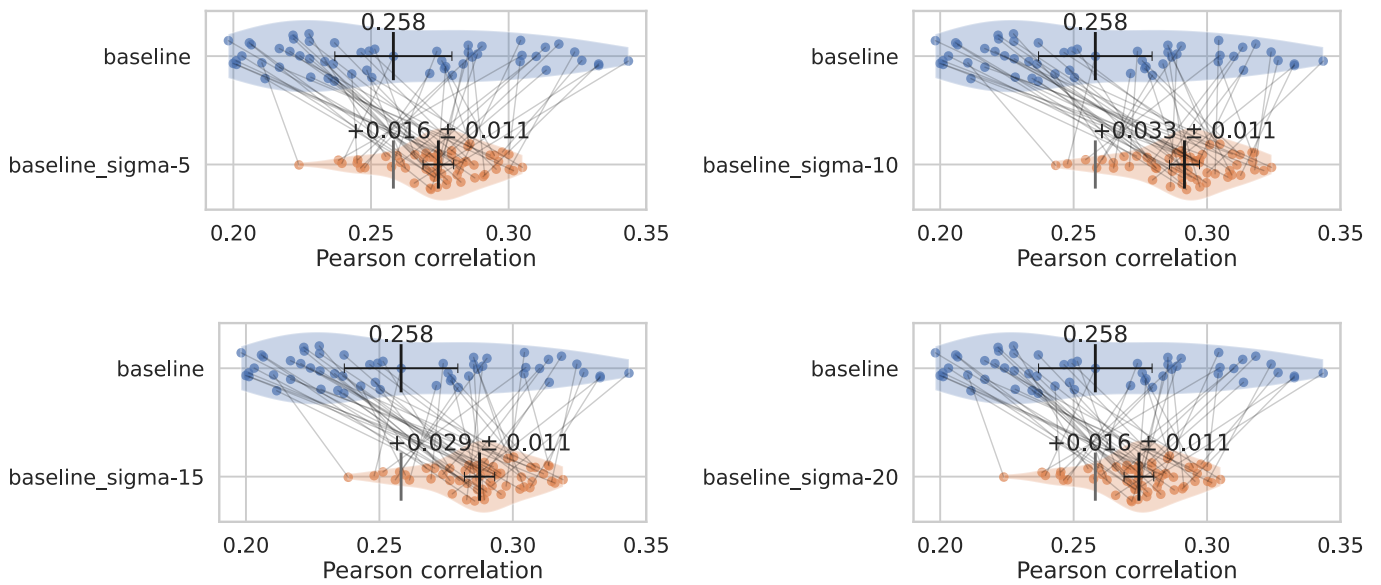


Figure 9.5: Comparison of gains in correlation after Gaussian blurring We compare correlation between subjects after the source subject’s functional data has been smoothed with a Gaussian kernel of standard deviation 5mm (top left), 10mm (top right), 15mm (bottom left) and 20mm (bottom right)

9.3 Control experiments

Controlling for smoothing effect increasing correlation Alignments computed with FUGW are not always vertex-to-vertex alignments. Indeed, a single vertex from the source subject s can be associated with many vertices in the target subject t . In fact, $P_i^{s,t}$ represents the relative importance of each match. The hyper-parameter ε controls the entropy of $P^{s,t}$, which is in direct link with the spread of vertices that we use as a measure for how many target vertices are matched with source vertices.

Since smoothing signal on the source subject can reduce noise and increase correlation to target data, we measure the correlation gain induced by applying a gaussian kernel to the source signal. This allows us to show that only a minor proportion of correlation gains induced by FUGW can come from this smoothing effect. Figure 9.5 shows this for kernels of 5mm, 10mm, 15mm and 20mm of standard deviation respectively. We see that correlation increases significantly less than when using FUGW (0.03 vs 0.12 correlation gain respectively). Moreover, one notices that even though correlation increases for pairs of subjects with a low initial correlation, it decreases for pairs with a high initial correlation. On the contrary, FUGW increases correlation for all pairs of subjects.

Different training sets yield comparable correlation gains While we leverage all IBC maps to derive our couplings, we show that the presented results hold when using a much smaller training dataset. In particular, we observe similar

correlation gains when using only the 57 maps of the Archi protocol for training (see Table 9.3). It takes about one hour per subject to acquire these maps, which we advocate is a reasonable amount of time to build a training set dedicated to align subjects within a given cohort (and possibly across cohorts). Finally, we train both FUGW and MSM with lower-dimensional versions of the previous datasets. To do so, given a pair of subjects (s, t) to be aligned, we fit a PCA on the left out subjects, project the data of subjects to be aligned on these components, and keep the first 20 components only. For both models, correlation gains remained unchanged.

More explicitly, we test the 4 following training sets:

- ALL-MATH: all contrast maps of IBC except contrasts from the Mathlang protocol (369 features per subject)
- ALL-MATH PCA: principal components fitted on ALL-MATH for all IBC subjects except s and t (20 features)
- ARCHI: all contrast maps from the Archi protocol of IBC (57 features)
- ARCHI PCA: principal components fitted on ARCHI for all subjects except s and t (20 features)

Training set	FUGW	MSM
ALL-MATH	0.12	0.01
ALL-MATH PCA	0.11	0.02
ARCHI	0.10	0.02
ARCHI PCA	0.11	0.01

Table 9.1: Gain in Pearson correlation of aligned contrast maps from the Mathlang protocol compared to the baseline The original correlation (baseline) is 0.258

Using naturalistic stimuli to derive alignments with FUGW This experiment's setup is similar to that of Experiment 1: Using training features to first derive OT couplings, we then use the latter to assess correlation gains between subjects's feature maps before and after subjects have been aligned. Naturalistic stimuli datasets include *Raiders of the Lost Ark*, short video clips and auditory stimuli from *The Little Prince* respectively adapted from [Bhattachali et al., 2019](#); [Haxby et al., 2011](#); [Nishimoto et al., 2011](#). Here, for each naturalistic dataset, we leverage work from [Richard, Gresele, Hyvarinen et al., 2020](#) to derive the first $m = 20$ components of a fitted shared response model. Share response models seek to find a common dictionary K of activation patterns across subjects $s \in S$ and to derive a mapping W^s with m orthogonal components that projects each individual's data onto this common space:

$$\begin{aligned} & \arg \min_{K, W^s} \sum_{s \in S} \|F^s - W^s K\|^2 \\ & \text{s.t. } (W^s)^T \cdot W^s = I_m \end{aligned}$$

These W^s are then used for alignment.

Results are reported in Table 9.2 and show that using datasets which are 20 times less time-consuming than that of Experiment 1 can already yield significant correlation gain on unseen task data.

Table 9.2: Acquisition time (AT) and correlation gain on the left hemisphere (CG) per training set (baseline correlation = 0.258)

Training set	Type	AT (min)	CG (Pearson)
All-MATH	tasks	2000	0.118
Clips	movie	100	0.017
Raiders	movie	115	0.046
The Little Prince	movie	100	0.009

Transporting myelin maps shows mild effect Leveraging transport plans computed using fMRI data from Experiment 1, we transport myelin maps – approximated through T1 / T2 ratio maps – from the source subject to the target subject. We compare the correlation of the unaligned source and target maps with the correlation of the transported and target maps. As illustrated in Figure 9.6, correlation gain is barely significant.

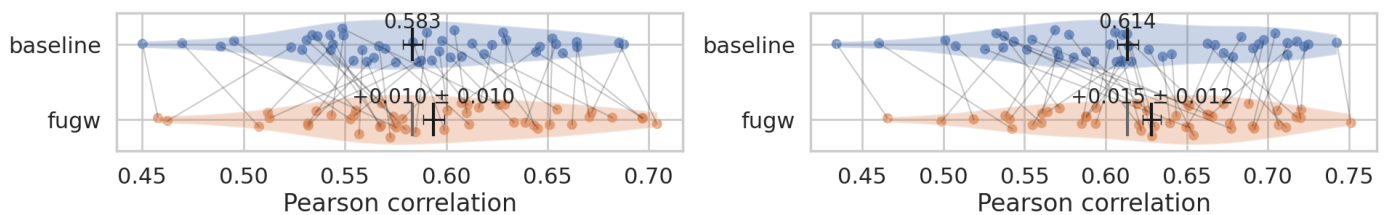


Figure 9.6: Comparison of gains in correlation after inter-subject alignment for myelin maps For each pair of source and target subjects of the IBC dataset, we compute the average Pearson correlation between myelin maps – approximated using T1/T2 ratios – for the left (left panel) and right (right panel) hemispheres. Correlation gains are not significant.

Before computing correlation between aforementioned maps, we discarded vertices located in the *cortical wall*, as they mostly contain spurious values. To do so, we set their value to the median of values of vertices which do not belong to the cortical wall. In order to determine which vertices belong to the *wall*, we used the Destrieux atlas [Destrieux et al., 2010](#).

Eventually, we advocate that little gain can be obtained when better aligning myelin maps, since they are already very stable across human subjects as shown in Figure 9.7.

9.4 Dataset description

The presented experiments rely on the Individual Brain Charting (IBC) dataset. A detailed description of the preprocessing pipeline of the IBC data is provided

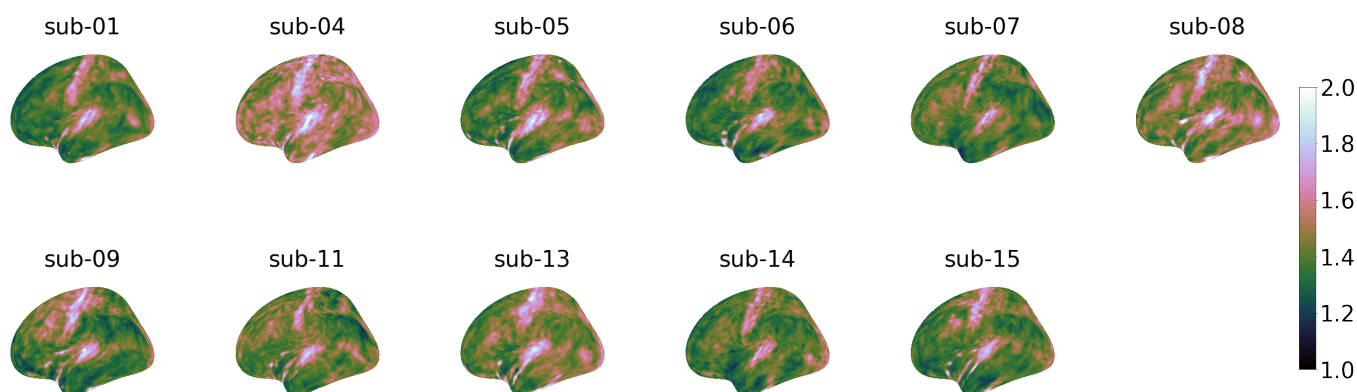


Figure 9.7: Myelin maps (approximated using T1/T2 ratio maps) are very consistent across IBC participants

in [Pinho et al., 2021](#). Raw data were preprocessed using *PyPreprocess* ².

All fMRI images, i.e. GE-EPI volumes, were collected twice with reversed phase-encoding directions, resulting in pairs of images with distortions going in opposite directions. Susceptibility-induced off-resonance field was estimated from the two Spin-Echo EPI volumes in reversed phase-encoding directions. The images were corrected based on the estimated deformation model. Details about the method can be found in [Andersson et al., 2003](#).

Further, the GE-EPI volumes were aligned to each other within every participant. A rigid-body transformation was employed, in which the average volume of all images was used as reference [Friston et al., 1995](#). The anatomical and motion-corrected fMRI images were given as input to *FreeSurfer* v6.0.0, in order to extract meshes of the tissue interfaces and the sampling of functional activation on these meshes, as described in [Van Essen et al., 2012](#). The corresponding maps were then resampled to the *fsaverage7* (high resolution, 163k nodes per hemisphere) and *fsaverage5* (low resolution, 10k nodes per hemisphere) templates of *FreeSurfer* [Fischl et al., 1999](#).

fMRI data were analyzed using the *General Linear Model*. Regressors of the model were designed to capture variations in BOLD response strictly following stimulus timing specifications. They were estimated through the convolution of boxcar functions, that represent per-condition stimulus occurrences, with the canonical *Hemodynamic Response Function* (HRF). To build such models, paradigm descriptors grouped in triplets (i.e. onset time, duration and trial type) according to BIDS Specification were determined from the log files' registries generated by the stimulus-delivery software. To account for small fluctuations in the latency of the HRF peak response, additional regressors were computed based on the convolution of the same task-conditions profile with the time derivative of the HRF. Nuisance regressors were also added to the design matrix in order to minimize the final residual error. To remove signal variance

²<https://github.com/neurospin/pypreprocess>

associated with spurious effects arising from movements, six temporal regressors were defined for the motion parameters. Further, the first five principal components of the signal, extracted from voxels showing the 5% highest variance, were also regressed to capture physiological noise [Behzadi et al., 2007](#).

In addition, a discrete-cosine basis was included for high-pass filtering ($cut-off = \frac{1}{128}$ Hz). Model specification was implemented using *Nilearn* v0.8.1 [Abraham et al., 2014](#), a Python library for statistical learning on neuroimaging data (<https://nilearn.github.io>).

In tables [9.3](#), [9.4](#) and [9.5](#), we give the explicit list of contrast and condition maps used for training, validation and testing, respectively.

Table 9.3: Training data This dataset comprises a wide variety of tasks: motor, visual, auditory, relational, linguistic, etc.

Training data	
Task	Condition / Contrast
ArchiEmotional	expression_control
ArchiEmotional	expression_gender
ArchiEmotional	expression_gender-control
ArchiEmotional	expression_intention
ArchiEmotional	expression_intention-control
ArchiEmotional	expression_intention-gender
ArchiEmotional	face_control
ArchiEmotional	face_gender
ArchiEmotional	face_gender-control
ArchiEmotional	face_trusty
ArchiEmotional	face_trusty-control
ArchiEmotional	face_trusty-gender
ArchiEmotional	trusty_and_intention-control
ArchiEmotional	trusty_and_intention-gender
ArchiSocial	false_belief-mechanistic
ArchiSocial	false_belief-mechanistic_audio
ArchiSocial	false_belief-mechanistic_video
ArchiSocial	false_belief_audio
ArchiSocial	false_belief_video
ArchiSocial	mechanistic_audio
ArchiSocial	mechanistic_video
ArchiSocial	non_speech_sound
ArchiSocial	speech-non_speech
ArchiSocial	speech_sound
ArchiSocial	triangle_mental
ArchiSocial	triangle_mental-random
ArchiSocial	triangle_random
ArchiSpatial	grasp-orientation
ArchiSpatial	hand-side
ArchiSpatial	object_grasp
ArchiSpatial	object_orientation
ArchiSpatial	rotation_hand
ArchiSpatial	rotation_side
ArchiSpatial	saccades
ArchiStandard	audio_computation
ArchiStandard	audio_left_button_press
ArchiStandard	audio_right_button_press
ArchiStandard	audio_sentence

Training data (next)

Task	Condition / Contrast
ArchiStandard	cognitive-motor
ArchiStandard	computation
ArchiStandard	computation-sentences
ArchiStandard	horizontal-vertical
ArchiStandard	horizontal_checkerboard
ArchiStandard	left-right_button_press
ArchiStandard	listening-reading
ArchiStandard	motor-cognitive
ArchiStandard	reading-checkerboard
ArchiStandard	reading-listening
ArchiStandard	right-left_button_press
ArchiStandard	sentences
ArchiStandard	sentences-computation
ArchiStandard	vertical-horizontal
ArchiStandard	vertical_checkerboard
ArchiStandard	video_computation
ArchiStandard	video_left_button_press
ArchiStandard	video_right_button_press
ArchiStandard	video_sentence
Attention	double_congruent
Attention	double_cue
Attention	double_incongruent
Attention	double_incongruent-double_congruent
Attention	incongruent-congruent
Attention	spatial_congruent
Attention	spatial_cue
Attention	spatial_cue-double_cue
Attention	spatial_incongruent
Attention	spatial_incongruent-spatial_congruent
Audi	alphabet
Audi	alphabet-silence
Audi	animals
Audi	animals-silence
Audi	cough
Audi	cough-silence
Audi	environment
Audi	environment-silence
Audi	human
Audi	human-silence
Audi	laugh

Training data (next)

Task	Condition / Contrast
Audi	laugh-silence
Audi	music
Audi	music-silence
Audi	reverse
Audi	reverse-silence
Audi	silence
Audi	speech
Audi	speech-silence
Audi	suomi
Audi	suomi-silence
Audi	tear
Audi	tear-silence
Audi	yawn
Audi	yawn-silence
Audio	animal
Audio	animal-others
Audio	animal-silence
Audio	mean-silence
Audio	music
Audio	music-others
Audio	music-silence
Audio	nature
Audio	nature-others
Audio	nature-silence
Audio	speech
Audio	speech-others
Audio	speech-silence
Audio	tool
Audio	tool-others
Audio	tool-silence
Audio	voice
Audio	voice-others
Audio	voice-silence
Bang	no_talk
Bang	talk
Bang	talk-no_talk
ColumbiaCards	gain
ColumbiaCards	loss
ColumbiaCards	num_loss_cards
Discount	amount

Training data (next)

Task	Condition / Contrast
Discount	delay
DotPatterns	correct_cue-incorrect_cue
DotPatterns	correct_cue_correct_probe
DotPatterns	correct_cue_incorrect_probe
DotPatterns	correct_cue_incorrect_probe-correct_cue_correct_probe
DotPatterns	correct_cue_incorrect_probe-incorrect_cue_correct_probe
DotPatterns	cue
DotPatterns	incorrect_cue_correct_probe
DotPatterns	incorrect_cue_incorrect_probe
DotPatterns	incorrect_cue_incorrect_probe-correct_cue_incorrect_probe
DotPatterns	incorrect_cue_incorrect_probe-incorrect_cue_correct_probe
DotPatterns	incorrect_probe-correct_probe
EmotionalPain	emotional-physical_pain
EmotionalPain	emotional_pain
EmotionalPain	physical_pain
Enumeration	enumeration_constant
Enumeration	enumeration_linear
Enumeration	enumeration_quadratic
Lec1	pseudoword
Lec1	pseudoword-random_string
Lec1	random_string
Lec1	word
Lec1	word-pseudoword
Lec1	word-random_string
Lec2	attend
Lec2	attend-unattend
Lec2	unattend
MCSE	high-low_salience
MCSE	high_salience_left
MCSE	high_salience_right
MCSE	low+high_salience
MCSE	low-high_salience
MCSE	low_salience_left
MCSE	low_salience_right
MCSE	salience_left-right
MCSE	salience_right-left
MTTNS	northside-southside_event
MTTNS	sn_after-before_event
MTTNS	sn_after_event
MTTNS	sn_all_event_response

Training data (next)

Task	Condition / Contrast
MTTNS	sn_all_space-time_cue
MTTNS	sn_all_space_cue
MTTNS	sn_all_time-space_cue
MTTNS	sn_all_time_cue
MTTNS	sn_average_event
MTTNS	sn_average_reference
MTTNS	sn_before-after_event
MTTNS	sn_before_event
MTTNS	sn_northside_event
MTTNS	sn_southside_event
MTTNS	sn_space-time_event
MTTNS	sn_space_event
MTTNS	sn_time-space_event
MTTNS	sn_time_event
MTTNS	southside-northside_event
MTTWE	eastside-westside_event
MTTWE	we_after-before_event
MTTWE	we_after_event
MTTWE	we_all_event_response
MTTWE	we_all_space-time_cue
MTTWE	we_all_space_cue
MTTWE	we_all_time-space_cue
MTTWE	we_all_time_cue
MTTWE	we_average_event
MTTWE	we_average_reference
MTTWE	we_before-after_event
MTTWE	we_before_event
MTTWE	we_eastside_event
MTTWE	we_space-time_event
MTTWE	we_space_event
MTTWE	we_time-space_event
MTTWE	we_time_event
MTTWE	we_westside_event
MTTWE	westside-eastside_event
MVEB	2_letters_different
MVEB	2_letters_different-same
MVEB	2_letters_same
MVEB	4_letters_different
MVEB	4_letters_different-same
MVEB	4_letters_same

Training data (next)

Task	Condition / Contrast
MVEB	6_letters_different
MVEB	6_letters_different-2_letters_different
MVEB	6_letters_different-same
MVEB	6_letters_same
MVEB	letter_occurrence_response
MVIS	2_dots-2_dots_control
MVIS	4_dots-4_dots_control
MVIS	6_dots-2_dots
MVIS	6_dots-6_dots_control
MVIS	dot_displacement_response
MVIS	dots-control
Moto	finger_left-fixation
Moto	finger_right-fixation
Moto	foot_left-fixation
Moto	foot_right-fixation
Moto	hand_left-fixation
Moto	hand_right-fixation
Moto	instructions
Moto	saccade-fixation
Moto	tongue-fixation
PainMovie	movie_mental
PainMovie	movie_mental-pain
PainMovie	movie_pain
Preference	face-others
Preference	food-others
Preference	house-others
Preference	painting-others
Preference	preference_constant
Preference	preference_linear
Preference	preference_quadratic
PreferenceFaces	face_constant
PreferenceFaces	face_linear
PreferenceFaces	face_quadratic
PreferenceFood	food_constant
PreferenceFood	food_linear
PreferenceFood	food_quadratic
PreferenceHouses	house_constant
PreferenceHouses	house_linear
PreferenceHouses	house_quadratic
PreferencePaintings	painting_constant

Training data (next)

Task	Condition / Contrast
PreferencePaintings	painting_linear
PreferencePaintings	painting_quadratic
RSVPLanguage	complex
RSVPLanguage	complex-consonant_string
RSVPLanguage	complex-simple
RSVPLanguage	consonant_string
RSVPLanguage	jabberwocky
RSVPLanguage	jabberwocky-consonant_string
RSVPLanguage	jabberwocky-pseudo
RSVPLanguage	probe
RSVPLanguage	pseudo-consonant_string
RSVPLanguage	pseudoword_list
RSVPLanguage	sentence-consonant_string
RSVPLanguage	sentence-jabberwocky
RSVPLanguage	sentence-pseudo
RSVPLanguage	sentence-word
RSVPLanguage	simple
RSVPLanguage	simple-consonant_string
RSVPLanguage	word-consonant_string
RSVPLanguage	word-pseudo
RSVPLanguage	word_list
SelectiveStopSignal	go_critical
SelectiveStopSignal	go_critical-stop
SelectiveStopSignal	go_noncritical
SelectiveStopSignal	go_noncritical-ignore
SelectiveStopSignal	ignore
SelectiveStopSignal	ignore-stop
SelectiveStopSignal	stop
SelectiveStopSignal	stop-ignore
Self	correct_rejection
Self	encode_other
Self	encode_self
Self	encode_self-other
Self	false_alarm
Self	instructions
Self	recognition_hit
Self	recognition_hit-correct_rejection
Self	recognition_other_hit
Self	recognition_self-other
Self	recognition_self_hit

Training data (next)

Task	Condition / Contrast
StopSignal	go
StopSignal	stop
StopSignal	stop-go
Stroop	congruent
Stroop	incongruent
Stroop	incongruent-congruent
TheoryOfMind	belief
TheoryOfMind	belief-photo
TheoryOfMind	photo
TwoByTwo	cue_switch-stay
TwoByTwo	cue_taskstay_cuestay
TwoByTwo	cue_taskstay_cueswitch
TwoByTwo	cue_taskswitch_cuestay
TwoByTwo	cue_taskswitch_cueswitch
TwoByTwo	stim_taskstay_cuestay
TwoByTwo	stim_taskstay_cueswitch
TwoByTwo	stim_taskswitch_cuestay
TwoByTwo	stim_taskswitch_cueswitch
TwoByTwo	task_switch-stay
VSTM	vstm_constant
VSTM	vstm_linear
VSTM	vstm_quadratic
Visu	animal
Visu	animal-scrambled
Visu	characters
Visu	characters-scrambled
Visu	face
Visu	face-scrambled
Visu	house
Visu	house-scrambled
Visu	pseudoword
Visu	pseudoword-scrambled
Visu	scene
Visu	scene-scrambled
Visu	scrambled
Visu	target_fruit
Visu	tool
Visu	tool-scrambled
WardAndAllport	ambiguous-unambiguous
WardAndAllport	intermediate-direct

Training data (next)	
Task	Condition / Contrast
WardAndAllport	move_ambiguous_direct
WardAndAllport	move_ambiguous_intermediate
WardAndAllport	move_unambiguous_direct
WardAndAllport	move_unambiguous_intermediate
WardAndAllport	planning_ambiguous_direct
WardAndAllport	planning_ambiguous_intermediate
WardAndAllport	planning_unambiguous_direct
WardAndAllport	planning_unambiguous_intermediate

Table 9.4: Validation data

Validation data	
Task	Condition / Contrast
HcpEmotion	face
HcpEmotion	face-shape
HcpEmotion	shape
HcpEmotion	shape-face
HcpGambling	punishment
HcpGambling	punishment-reward
HcpGambling	reward
HcpGambling	reward-punishment
HcpLanguage	math
HcpLanguage	math-story
HcpLanguage	story
HcpLanguage	story-math
HcpMotor	cue
HcpMotor	left_foot
HcpMotor	left_foot-avg
HcpMotor	left_hand
HcpMotor	left_hand-avg
HcpMotor	right_foot
HcpMotor	right_foot-avg
HcpMotor	right_hand
HcpMotor	right_hand-avg
HcpMotor	tongue
HcpMotor	tongue-avg
HcpRelational	match
HcpRelational	relational
HcpRelational	relational-match
HcpSocial	mental
HcpSocial	mental-random
HcpSocial	random
HcpWm	0back-2back
HcpWm	0back_body
HcpWm	0back_face
HcpWm	0back_place
HcpWm	0back_tools
HcpWm	2back-0back
HcpWm	2back_body
HcpWm	2back_face
HcpWm	2back_place

Validation data (next)

Task	Condition / Contrast
HcpWm	2back_tools
HcpWm	body-avg
HcpWm	face-avg
HcpWm	place-avg
HcpWm	tools-avg

Table 9.5: Test data

Test data	
Task	Condition / Contrast
MathLanguage	arithmetic_fact-othermath
MathLanguage	arithmetic_fact_auditory
MathLanguage	arithmetic_fact_visual
MathLanguage	arithmetic_principle-othermath
MathLanguage	arithmetic_principle_auditory
MathLanguage	arithmetic_principle_visual
MathLanguage	auditory-visual
MathLanguage	colorlessg-wordlist
MathLanguage	colorlessg_auditory
MathLanguage	colorlessg_visual
MathLanguage	context-general
MathLanguage	context-theory_of_mind
MathLanguage	context_auditory
MathLanguage	context_visual
MathLanguage	general-colorlessg
MathLanguage	general_auditory
MathLanguage	general_visual
MathLanguage	geometry-othermath
MathLanguage	geometry_fact_auditory
MathLanguage	geometry_fact_visual
MathLanguage	math-nonmath
MathLanguage	nonmath-math
MathLanguage	theory_of_mind-context
MathLanguage	theory_of_mind-general
MathLanguage	theory_of_mind_and_context-general
MathLanguage	theory_of_mind_auditory
MathLanguage	theory_of_mind_visual
MathLanguage	visual-auditory
MathLanguage	wordlist_auditory
MathLanguage	wordlist_visual

Chapter 10

Supplementary material for Chapter 4 (FUGW extensions)

10.1 Optimizing convergence speed

Figures [10.1](#) and [10.2](#) show the convergence speed of the three solvers for the FUGW problem when fitting using 100 and 220 IBC contrast maps of the Archi and HCP protocols respectively.

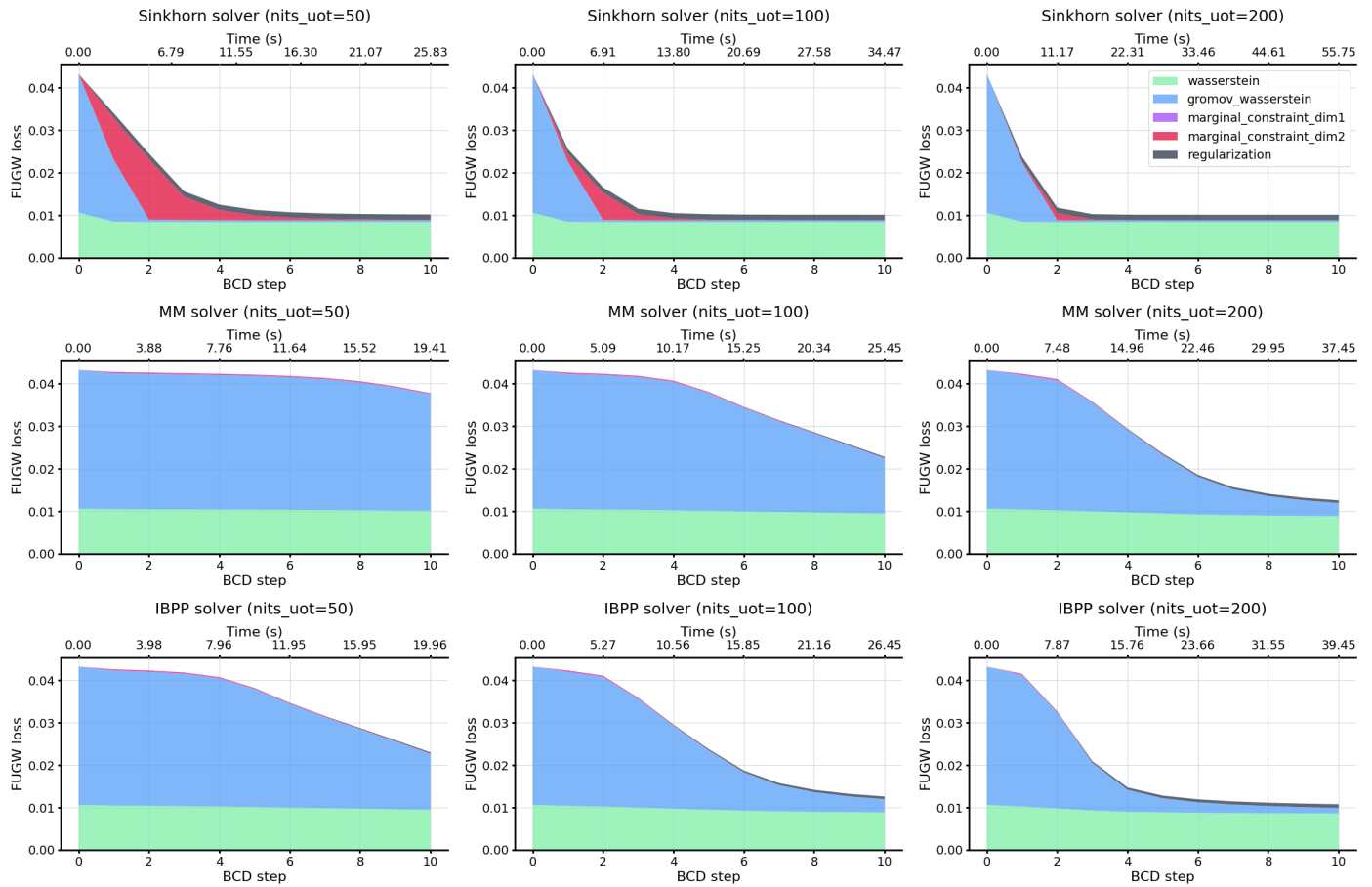


Figure 10.1: Convergence speed comparison of three solvers for the FUGW problem when fitting using 100 IBC contrast maps of the Archi and HCP protocols. Iterations of the MM and IBPP algorithms (second and third rows respectively) are faster than those of the Sinkhorn algorithm (first row). However, their convergence is slower. With our data of interest, we see that 100 iterations of Sinkhorn's algorithm ran 5 to 6 times (one time for each BCD iteration) are enough to reach convergence in about 20 seconds on a single Tesla V100-DGXS-32GB.

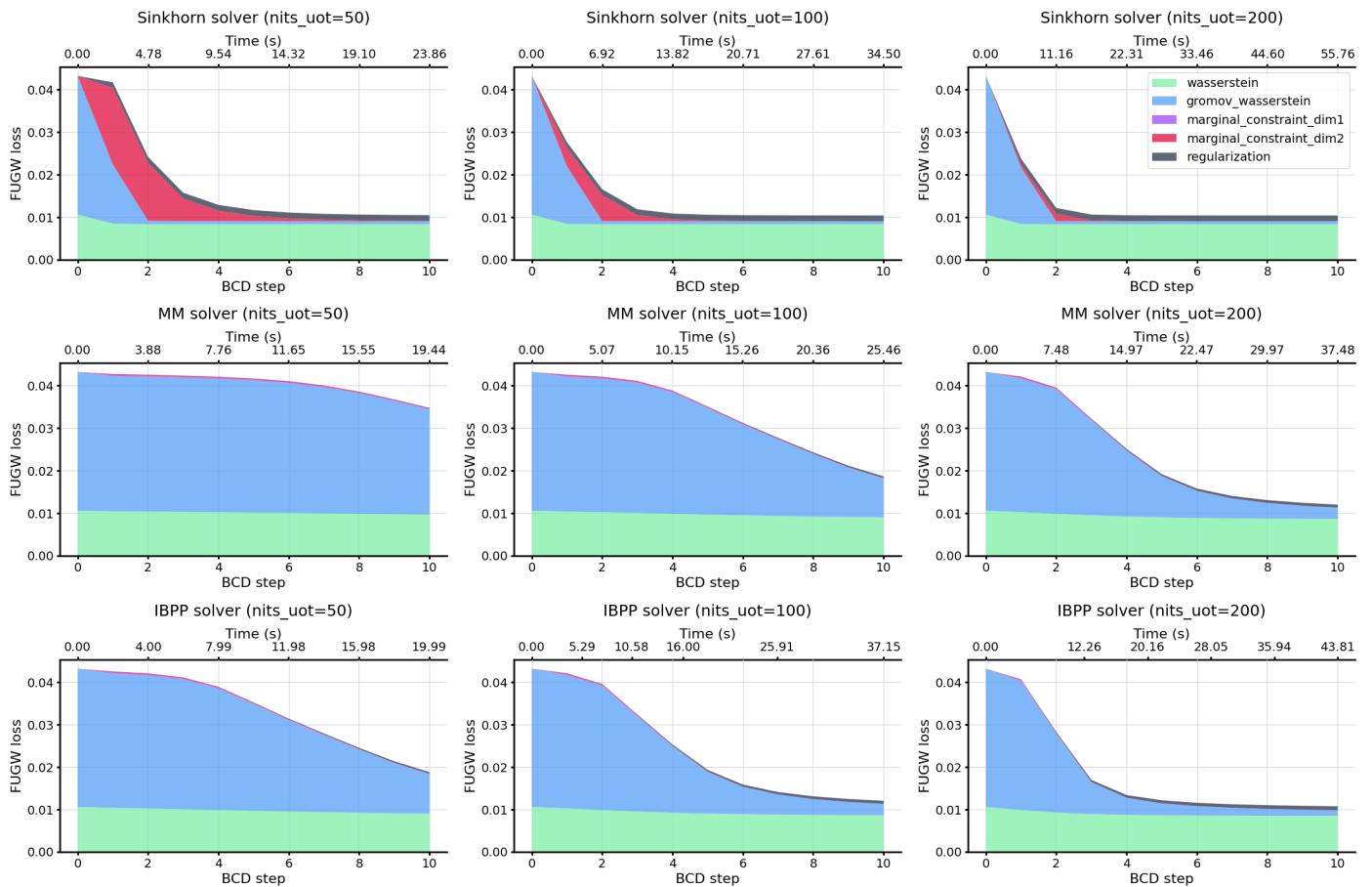


Figure 10.2: Convergence speed comparison of three solvers for the FUGW problem when fitting using 220 IBC contrast maps of the Archi, HCP, RSV-PLanguage, Preference, MathLanguage, FaceBody and Emotion protocols. Iterations of the MM and IBPP algorithms (second and third rows respectively) are faster than those of the Sinkhorn algorithm (first row). However, their convergence is slower. With our data of interest, we see that 100 iterations of Sinkhorn’s algorithm ran 5 to 6 times (one time for each BCD iteration) are enough to reach convergence in about 20 seconds on a single Tesla V100-DGX5-32GB.

Chapter 11

Supplementary material for Chapter 6 (Inter-subject visual decoding)

11.1 Data pre-processing

We strive to minimally preprocess acquired BOLD signal. To this end, we detrend acquired BOLD signal (i.e. we remove cosine drifts) and finally standardise voxels' timecourses for each run, as shown in Figure 11.1.

Moreover, when decoding the latent representation of a given image, we use brain volumes which have been acquired after the image's onset. Figure 11.2 illustrates this idea, and introduces the concepts of *window size* (i.e. the number of brain volumes we use) and *lag* (i.e. the time difference between the onset of the image to be decoded and the first brain volume used to decode it). Values for both of these hyper-parameters were obtained through a 5-fold cross-validated grid search over samples of the training set. We report these results in Figure 11.3.

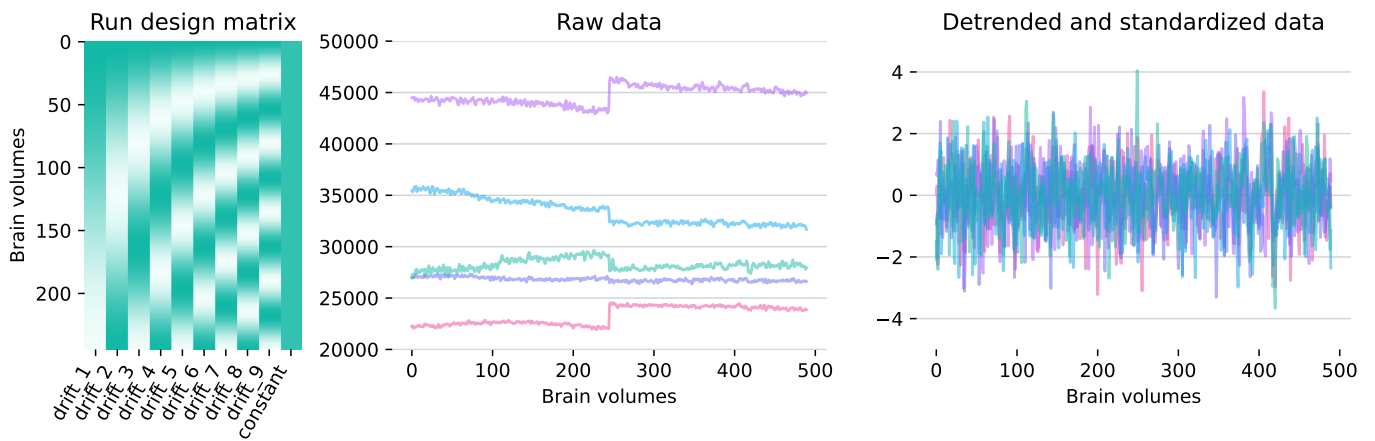


Figure 11.1: Pre-processing of the Wen 2017 dataset For each participant and each run, in each vertex, we regress out parts of the signal which can be linearly explained by the design matrix represented on the left, which models cosine drifts of the BOLD signal. The two graphs to the right show time-courses in 5 vertices across 2 different runs before (left) and after (right) they have been pre-processed.

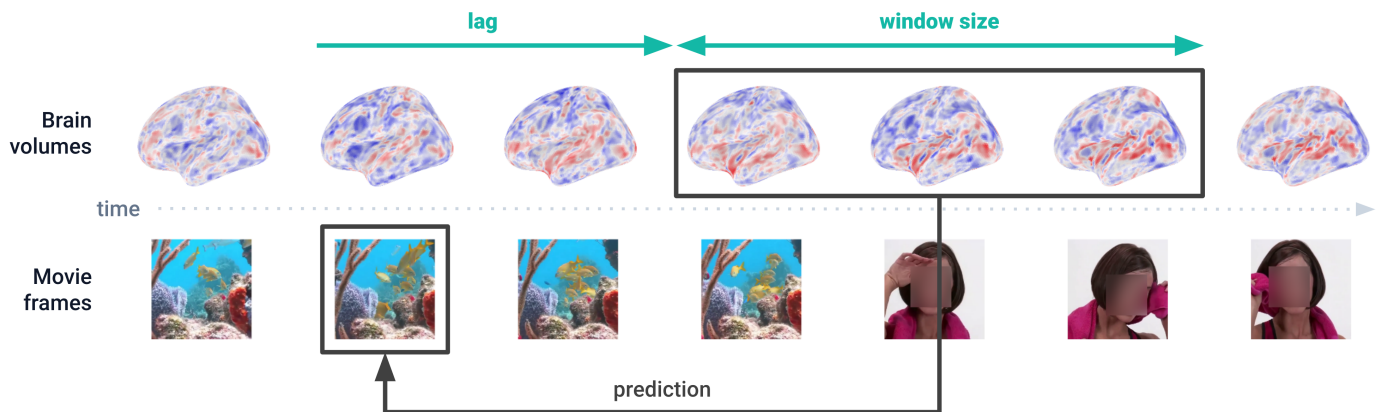


Figure 11.2: Lag and window size In order to decode a movie frame which was seen at time t , one can use brain volumes which were acquired further in time. This delay is referred to as the *lag*. Moreover, one can use several brain volumes to decode a given movie frame. The number of brain volumes used is called the *window size*. Images featuring human faces were blurred.

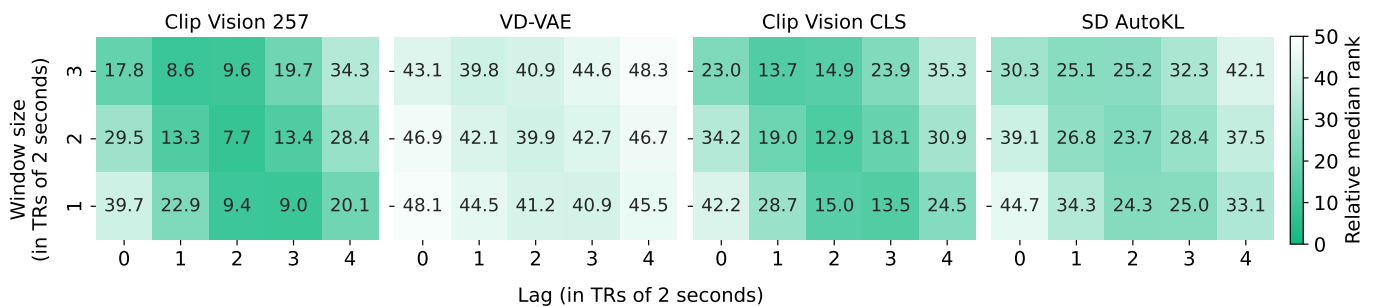


Figure 11.3: Relative median rank ↓ of predicted latents averaged across participants for various time lags and window sizes

11.2 Retrieving images using predicted latent representations

Predicted latent representations can be compared to that of images in a retrieval set. In Figure 11.4, for each image shown to the participant during the test phase, we print the five images from the retrieval set whose latent representation is the closest to predicted latents. We see that semantics are often preserved.

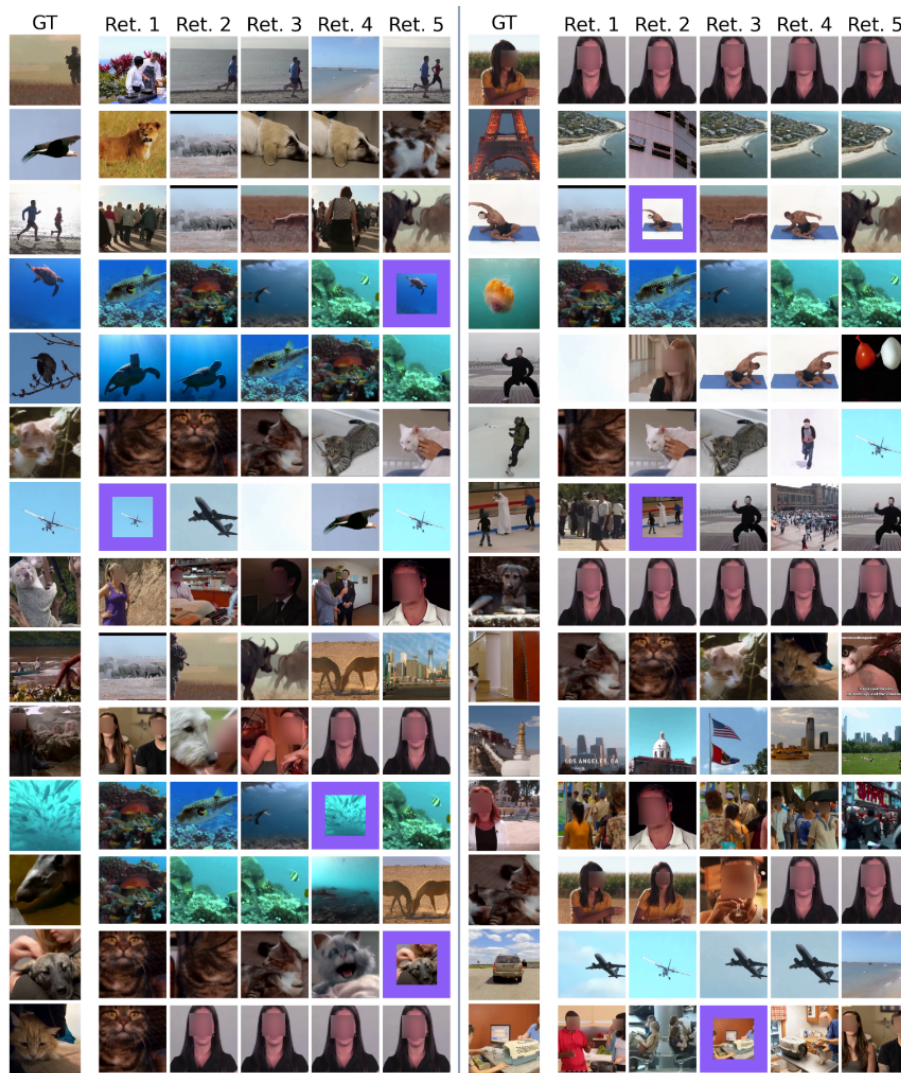


Figure 11.4: Image retrievals using predicted latent representations of CLIP 257×768 latents

We use a model fitted on Subject 2 (S2) from the Wen 2017 dataset and predict the latent representation of unseen videos (test set). Ground truth (GT) images featured within the first 5 retrieved (Ret.) images are indicated with a bold purple border. In a given row, images which appear similar across columns are actually different frames of the same video clip. Images featuring human faces were blurred.

11.3 Results for every combination of reference participant and left-out participant

Figure 11.5 is a copy of Figure 6.3 from the main pages of this paper. It illustrates the main effects reported in our study, namely that (1) functional alignment yields better performance than anatomical alignment when transferring a semantic decoder to left-out individuals, (2) it is possible to train such decoders on multiple participants and (3) this last setup works best when participants are aligned.

Figure 11.5 only shows these results when participant 2 of the Wen 2017 dataset is used as the reference participant. Therefore, we add Figures 11.6 and 11.7, which illustrate that all results hold regardless of what participant of the cohort is used as reference or left-out participant.

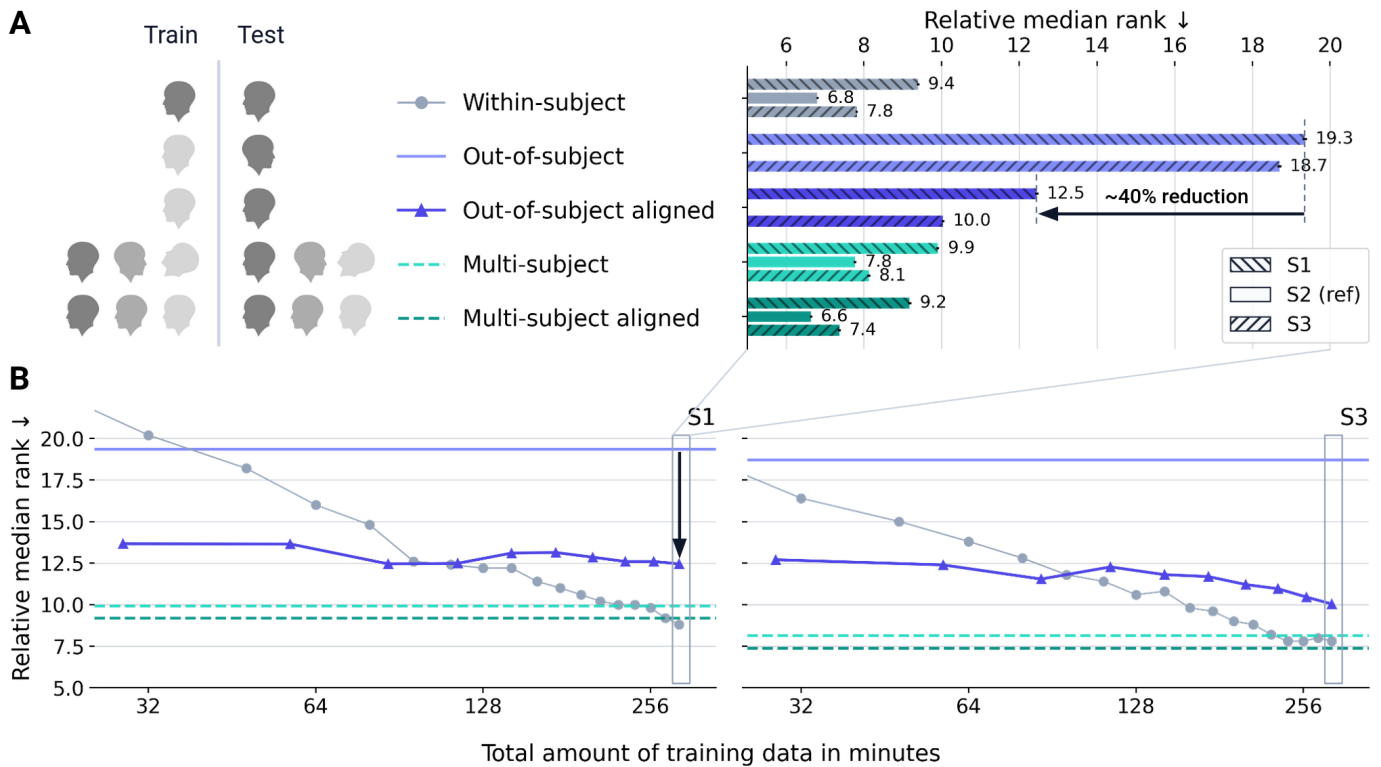


Figure 11.5: Effects of functional alignment on multi-subject and out-of-subject setups using participant 2 as the reference participant We report relative median rank ↓ in all setups described in section 6.2.3 for CLIP 257×768 . In all *aligned* cases, S1 and S3 were aligned onto S2. In all *out-of-subject* cases, we test S1 and S3 onto a decoder trained on S2. In all *multi-subject* cases, the decoder was trained on all data from all 3 participants. **A**. In this panel, all models (alignment and decoding) were trained on all available training data. Results for other latent types are available in Figure 11.8. **B**. In left-out S1 and S3, decoding performance is much better when using functional alignment to S2 (solid dark purple) than when using anatomical alignment only (solid pale purple). Performance increases slightly as the amount of data used to align participants grows, but does not always reach levels that can be achieved with a single-participant model fitted in left-out participants (solid pale gray dots) when a lot of training data is available. Training a model on multiple participants yields good performance in all 3 participants (dashed pale teal) which can be further improved by using functional alignment (dashed dark teal). Results for other latent types are available in Figure 11.9.

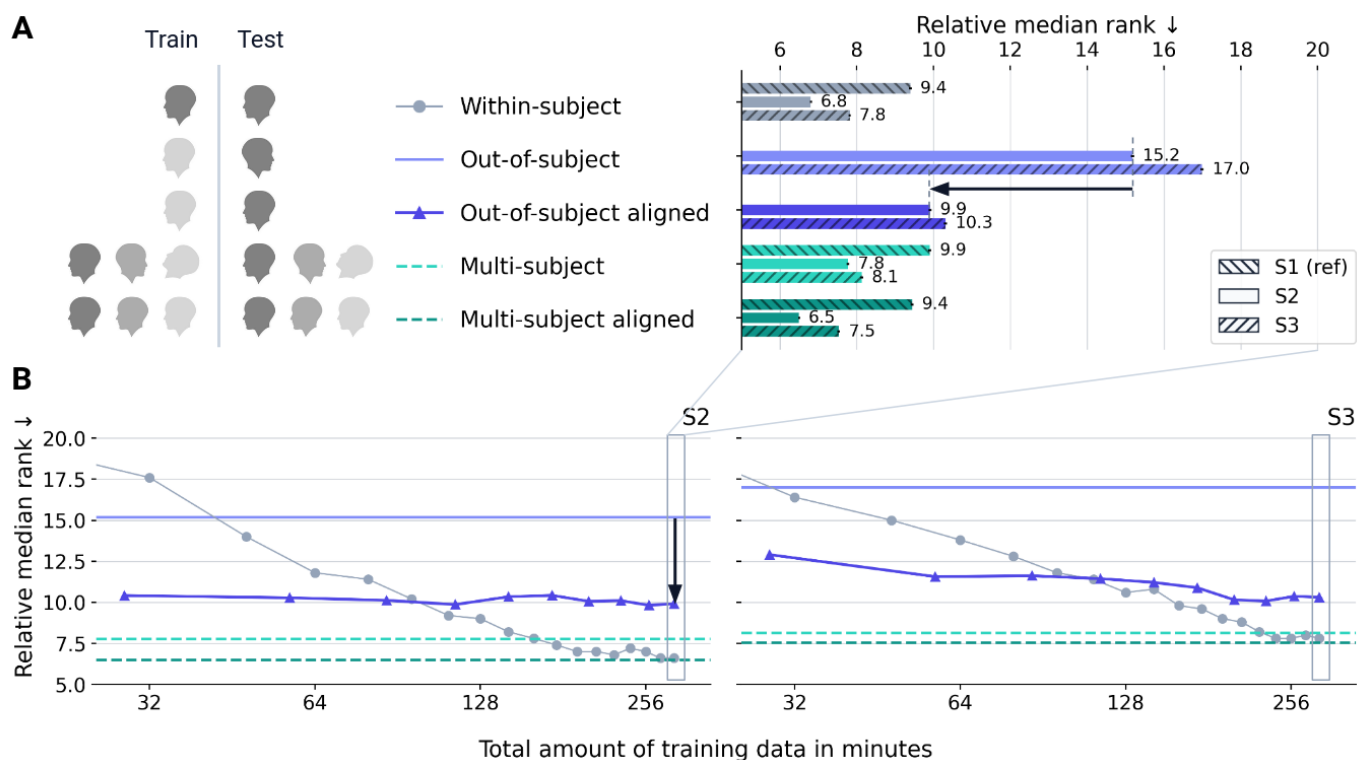


Figure 11.6: Effects of functional alignment on multi-subject and out-of-subject setups using participant 1 as the reference participant

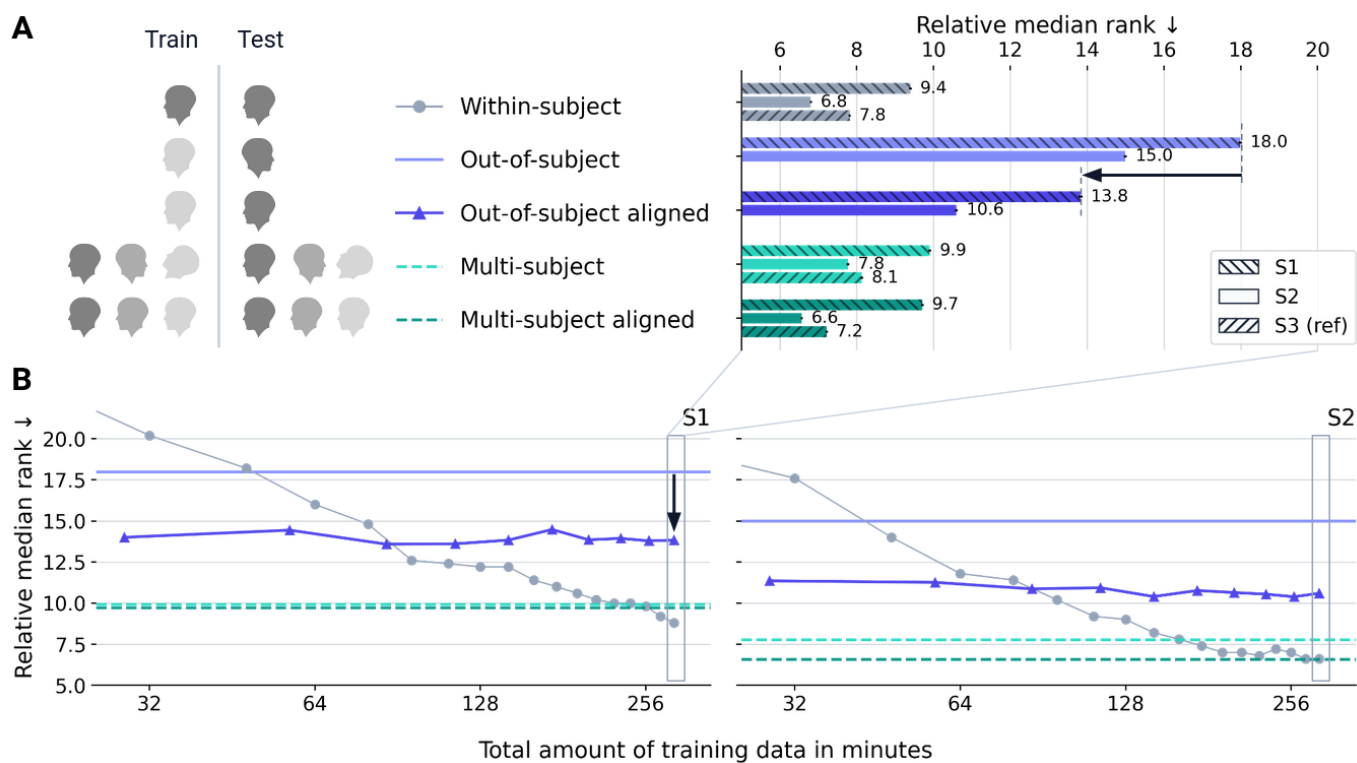


Figure 11.7: Effects of functional alignment on multi-subject and out-of-subject setups using participant 3 as the reference participant

11.4 Results for every type of latent representation

In this section, we extend the claims made in Figures 6.3.A, 6.3.B and 6.5 by showing that these results hold for other latent representations, namely VD-VAE, CLIP CLS and AutoKL. Figures 11.8, 11.9 and 11.10 extend Figures 6.3.A, 6.3.B and 6.5 respectively, showing that observed effects are present regardless of the chosen latent representation. All of these figures were obtained using the Wen 2017 dataset.

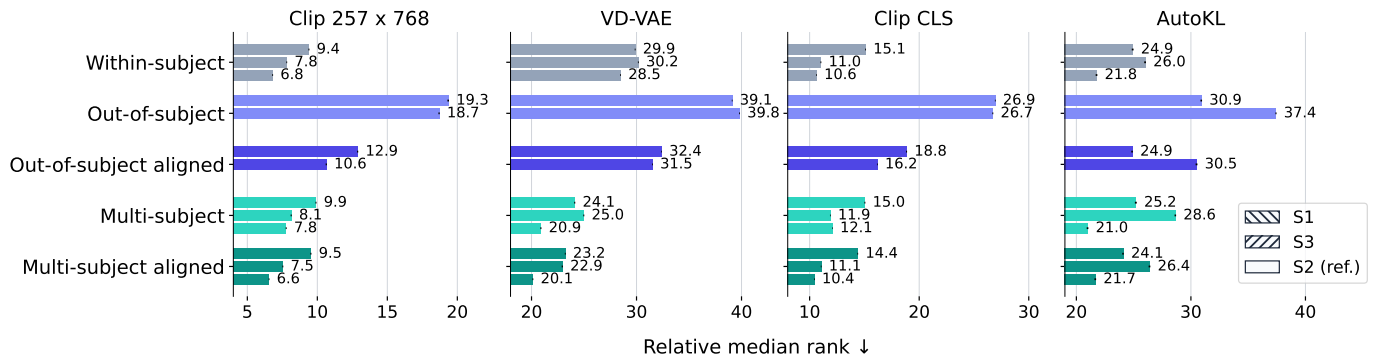


Figure 11.8: Effects of alignment For any type of latent representation, out-of-subject decoding performance, measured through relative median rank ↓, greatly improves when participants are functionally aligned. Training decoders on multiple participants also works better when participants are aligned. These results were averaged across 50 retrieval sets ; all these metrics are reported with a standard error of the mean (SEM) smaller than 0.01.

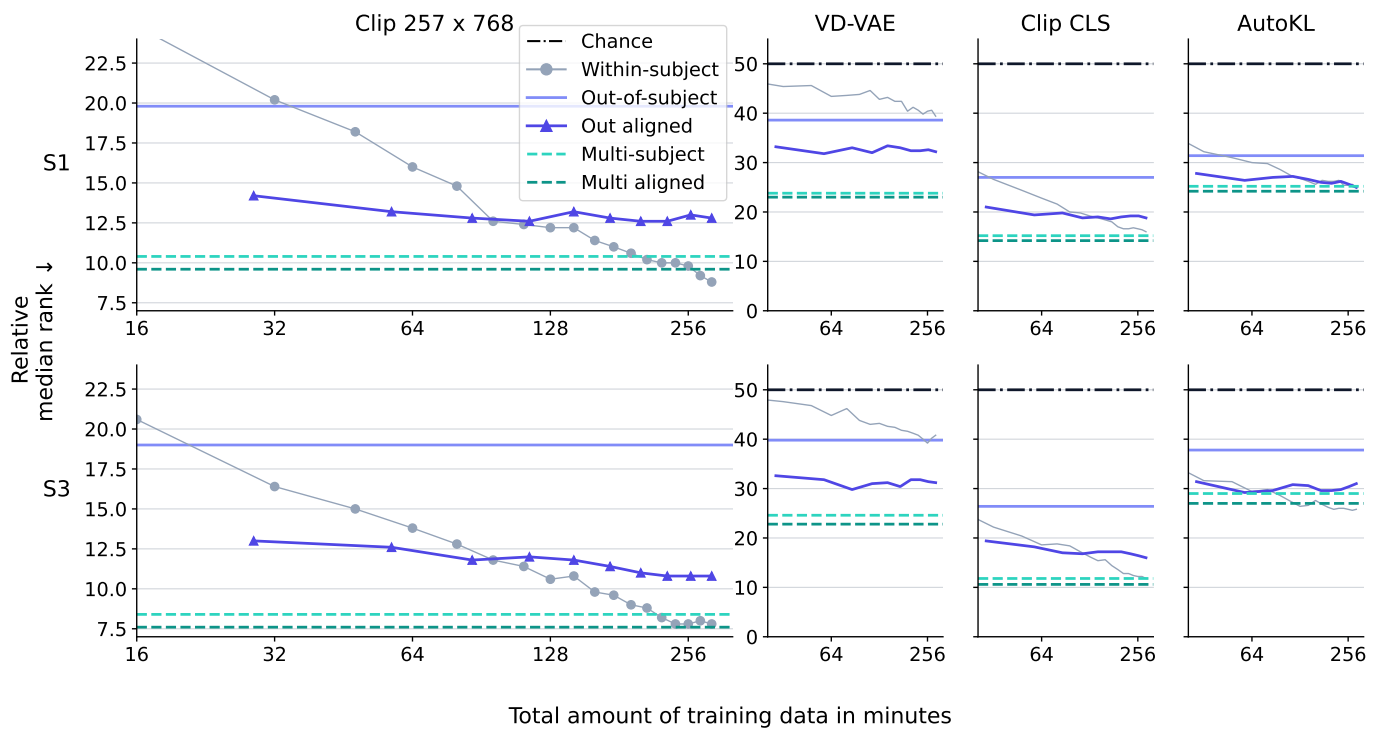


Figure 11.9: Performance increases slightly with more alignment data For any type of latent representation, out-of-subject decoding performance greatly increases with functional alignment even in low data regimes. In high data regimes, out-of-subject decoding does not work as well as fitting single-subject or multi-subject models.

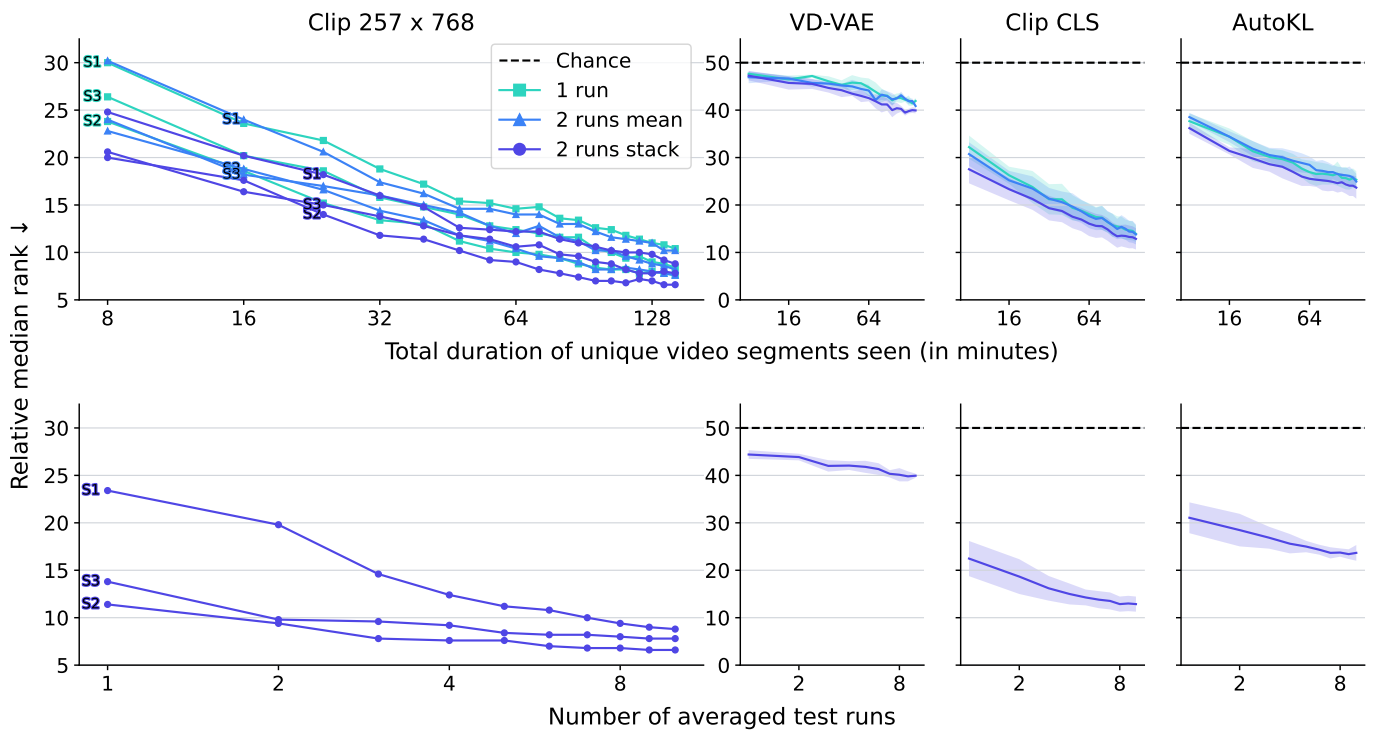


Figure 11.10: Scaling studies for all latents For any type of latent representation, decoding performance increases linearly with exponentially more data. It also seems that, when acquiring data at 3T or more, not repeating stimuli yields the best results. At test time, although repeating stimuli allows to get better metrics, retrieval performance with only one repetition is already reasonable in 2 out of 3 participants of the Wen 2017 dataset.

11.5 Decoding results for all setups

On top of experiments reported in the main pages of this paper, we have tested a lot of different training and tests sets. In this section, we report the Relative Median Rank \downarrow for all 97 training sets and all 63 test sets, and CLIP latent representations. Training sets include all possible single-subject, multi-subject unaligned and multi-subject aligned cases. Test sets include all possible with-subject, left-out unaligned and left-out aligned cases. Every time, all available training sessions are used for training the decoder. However, we vary the amount of data used to train the alignments, for both training and test sets.

We report detailed results for CLIP 257×768 and CLIP CLS in Figures [11.11](#) and [11.12](#) respectively.

In particular, these figures report combinations our setups of interest which were not mentioned in the main text. We find of particular interest the multi-subject out-of-subject setup, in which the decoder has been trained on two (un)aligned participants and tested on a third (un)aligned one.

11.6 Replication on the Natural Scenes Dataset

We replicate our main experiment using data from the Natural Scenes Dataset [Allen et al., 2022]. This dataset comprises 8 participants who each see 10 000 images 3 times, thus leading to a total of 30 000 trials per participant. For each participant, this data is acquired in 40 sessions of 60 minutes each. For each participant, there is a total of 1 000 images which are shared with other individuals - i.e. other individuals will see them too - and 9 000 which are exclusive - i.e. other individuals will not see them. We sub-selected all participants who had completed all 30 000 trials, namely participants 1, 2, 5, and 7. For each selected participant, we split their 30 000 trials in two sets: all exclusive images are grouped in the *decoding set* and all shared images are grouped in the *alignment set*. We further split the decoding set into disjoint sets of images for training and testing individual decoders, whose performance is reported in Table 11.1. Alignment sets are used to compute functional alignments between individuals. Besides, we used pre-computed beta coefficients computed with GLM denoise on *fsaverage7* [Fischl, 2012] and openly available online. We down-sampled this data to *fsaverage5* - which simply amounts to keeping only the first 10 242 array elements in each hemisphere.

Eventually, we show that decoders tested on left-out individuals work consistently and significantly better when left-out participants are functionally aligned rather than simply anatomically aligned to the reference participant, as reported in Table 11.2.

	CLIP 257 × 768		VD-VAE		CLIP CLS		AutoKL	
	MR	Acc	MR	Acc	MR	Acc	MR	Acc
S1	3.6	26.6	23.0	4.6	4.6	19.1	30.5	1.8
S2	6.0	17.6	22.1	4.4	6.9	13.8	33.6	1.5
S5	4.6	19.9	26.0	3.7	4.3	19.7	31.5	2.5
S7	4.0	24.4	23.4	5.4	5.5	18.5	24.5	4.3

Table 11.1: Within-subject metrics for all NSD participants and all latent types on the test set Reported metrics are relative median rank ↓ (MR) of retrieval on a set of 500 samples, top-5 accuracy % ↑ (Acc) of retrieval on a set of 500 samples. Chance level is at 50.0 and 1.0 for these metrics respectively. These results were averaged across 50 retrieval sets, hence results are reported with a standard error of the mean (SEM) smaller than 0.01.

Reference	Left-out	CLIP 257×768		VD-VAE		CLIP CLS		AutoKL	
		A	A+F	A	A+F	A	A+F	A	A+F
S1	S2	20.9	10.7	35.2	24.5	28.3	13.4	42.5	37.1
	S5	32.2	13.6	43.1	32.0	33.4	12.5	43.9	37.7
	S7	33.5	14.7	40.8	30.7	36.1	17.1	45.0	37.0
S2	S1	18.3	10.4	37.0	30.4	24.4	14.4	35.9	37.1
	S5	29.3	12.0	40.3	34.1	30.8	11.2	42.4	39.6
	S7	27.9	14.6	40.2	32.8	32.1	17.9	41.8	38.9
S5	S1	29.2	13.0	43.4	34.4	33.8	14.4	37.8	33.8
	S2	26.2	9.8	40.1	30.1	31.4	11.2	36.0	36.2
	S7	29.8	14.1	41.7	32.0	34.4	17.3	40.0	34.9
S7	S1	27.7	7.7	43.1	30.1	32.3	11.5	40.6	26.5
	S2	23.4	8.6	38.9	26.4	28.6	13.2	37.8	29.1
	S5	27.5	8.8	41.6	31.2	30.4	9.8	43.4	30.9

Table 11.2: Across-subject metrics for all NSD participants and all latent types on the test set We report the decoding performance of decoders trained on a reference participant and tested on a left-out participant who was anatomically aligned (A) or functionally aligned (A+F). The reported metric is the relative median rank \downarrow (MR) of retrieval on a set of 500 samples. These results were averaged across 50 retrieval sets, hence results are reported with a standard error of the mean (SEM) smaller than 0.01. One sees that functionally aligned data is always better decoded than anatomically aligned data. In particular, when S7 as the reference subject, functional alignment helps divide the median rank by 3 for CLIP latents.

Chapter 12

Supplementary material for Chapter 7 (Inter-species visual decoding)

12.1 Additional examples of predictions

We report random examples of predictions for both human and non-human subjects in Figures 12.1 and 12.2.

In Figure 12.1, the brain decoder was trained on all IBC participants. All participants were aligned on a reference participant using Clips-Train data. The decoder was trained on the Clips-Train data of all aligned participants. It was tested on the Clips-Valid data of the reference participant, here sub-04.

In Figure 12.1, the brain decoder was trained on all IBC participants. All participants were aligned on a reference participant using Clips-Train, Clips-Valid, and the first two segments of Monkey Kingdom data. The decoder was trained on the Clips-Train, Clips-Valid, and the first two segments of Monkey Kingdom data for all aligned participants. It was tested on the last two segments of Monkey Kingdom in non-human primates, here Luce. Luce was previously aligned to the reference participant, here sub-04, using the first two segments of Monkey Kingdom data.



Seen image	Caption generated from ground truth latents	Retrieved images	Caption generated from predicted latents
	A sea slug with a long beak and a long beak sitting in the water.		A large group of people walking around a parking lot.
	A red and white fire hydrant on a snowy surface.		A small island with a few animals on it.
	A clock that is on the side of a building.		A large, dark, yellow, and white clock tower.
	Two young boys in a field with balloons.		A man is walking on a cement walkway.
	A man is laying on a bed with a tie.		A man with a pair of fake eyes on his head.
	A beach with a large blue surfboard on it.		A large body of water with a lot of trees.
	A blurry image of a person in the dark.		A man is standing on a ledge with a toy.
	A polar bear is sitting in the middle of a body of water.		A large body of water with a lot of trees.
	A man with a ring on his finger in a black and white photo.		A man in a white shirt and a black and white striped tie.
	A zebra is standing on a rail in a black and white photo.		A group of people walking around a pile of rocks.
	A dog looking at a woman in a kitchen.		A young man in a black shirt and tie.
	A group of fish swimming in a body of water.		A large white bird sitting in the middle of a field.
	A close up of a glass of wine.		A view of a city with a lot of trees.
	A dog is looking at a figurine of a person.		A man wearing a white dress holding a small animal.

Figure 12.1: Random examples of predictions for a brain decoder trained and tested on human participants (Left) The ground truth image and caption generated using the ground truth latents. (Right) The first five retrieved images and the caption generated using the predicted latents.

Seen image	Caption generated from ground truth latents	Retrieved images	Caption generated from predicted latents
	A flock of birds flying over a forest.		A bunch of photos of a park with trees and bushes.
	A large stone structure with a clock on it.		A man with a large face sitting on a stage.
	A small animal sleeping in a tree filled with leaves.		A small bird is standing in the dark.
	A large tree with a large sun on top.		A man walking down a walkway next to a river.
	A baby monkey is wrapped in a tree branch.		A black and white photo of a plant in the distance.
	A couple of animals that are standing in the grass.		Two colorful kites flying in the air.
	A group of monkeys and a baby monkey in a forest.		A man is bending over to grab a dalmatian.
	A mother and baby monkey are wrapped in a tree.		A blurry photograph of a tree in the dark.
	A group of monkeys and a baby monkey hanging out together.		A view of a forest with a lot of trees.
	Two monkeys are touching each other on a ledge.		A man is on a table with a bunch of people.
	A monkey with a blonde hair on his head.		A man climbing a ladder to a large blue and white plane.
	A sky filled with clouds and a mountain.		A group of animals standing on top of a lush green field.
	A black bear climbing a tree with a branch.		A view of a bunch of trees and a building.
	A group of three monkeys walking through a forest.		A man sitting on a chair with a large brown elephant.
	A small brown animal with a green head.		A blurry photo of a tree in the dark.

Figure 12.2: Examples for predictions for a brain decoder trained on human participants and tested on a non-human primate (Left) The ground truth image and caption generated using the ground truth latents. (Right) The first five retrieved images and the caption generated using the predicted latents.

In Figure 12.3, we recall the decoding experiment for the Monkey Kingdom dataset. Namely, we train eight brain decoders. Each decoder is trained on all IBC subjects. All participants are aligned onto a reference participant using the Clips-Train, Clips-Valid, and the first two segments of Monkey Kingdom data. The same stimuli (Clips-Train, Clips-Valid, and the first two segments of Monkey Kingdom data) are used to train the decoder. We then test each decoder on the fourth and fifth segments of Monkey Kingdom in a left-out aligned macaque. Macaques were aligned to humans using the first two segments of Monkey Kingdom. We leave the third segment out to ensure no continuity between the third and fourth segments.

In Figure 12.4, we present the results for human participants in the same setup, except that Clips-Train was used to compute the functional alignments used at test time. We see that performance improves, suggesting that the Monkey Kingdom data does not provide enough information to compute functional alignments as accurate as those obtained with the Clips-Train data.

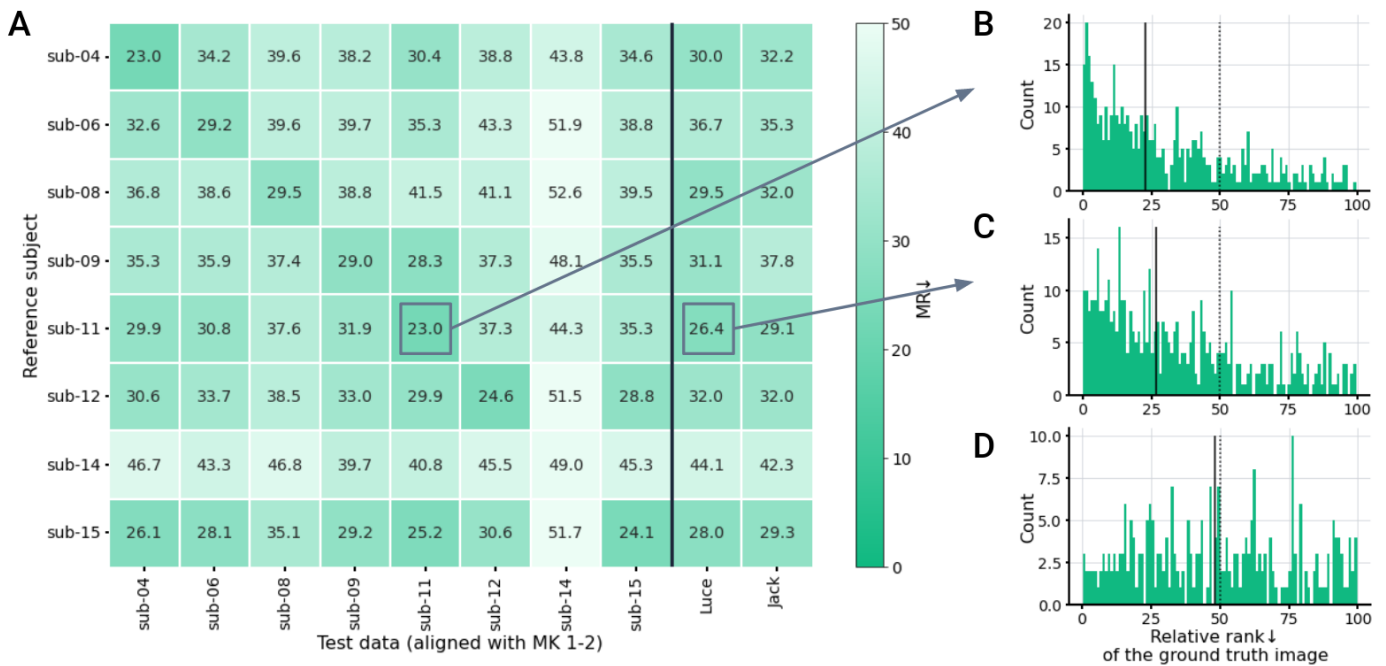


Figure 12.3: Brain decoders trained in human participants transfer to non-human participants **A.** Median relative rank for all 8 participants of IBC and 2 macaques (x-axis) when tested on a brain decoder trained on all IBC participants aligned onto a reference individual (y-axis). **B.** Distribution of the relative rank of the ground truth image in a human participant. **C.** Distribution of the relative rank of the ground truth image in a non-human primate. **D.** Distribution of the relative rank of the ground truth image when predictions for the non-human primate are randomly shuffled.

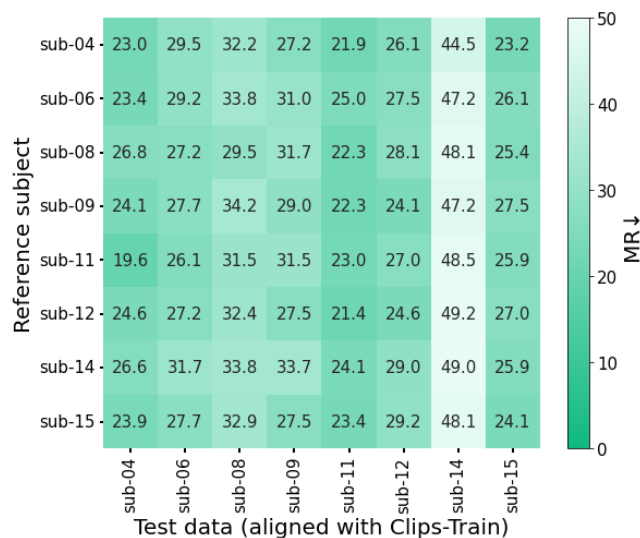


Figure 12.4: Control experiment for the Monkey Kingdom dataset Median relative rank for all 8 participants of the IBC dataset tested on 8 brain decoders. The difference with results reported in Figure 12.3 is that the functional alignments used at test time were computed using the Clips-Train data instead of the first two segments of Monkey Kingdom.

Bibliography

- Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8. <https://www.frontiersin.org/article/10.3389/fninf.2014.00014>
- Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, 25(1), 116–126. <https://doi.org/10.1038/s41593-021-00962-x>
- Al-Wasity, S., Vogt, S., Vuckovic, A., & Pollick, F. E. (2020). Hyperalignment of motor cortical areas based on motor imagery during action observation. *Sci Rep*, 10(1), 5362.
- Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, 113(18), 4909–4917. <https://doi.org/10.1073/pnas.1603205113>
- Andersson, J. L., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *Neuroimage*, 20(2), 870–888. [http://doi.org/10.1016/S1053-8119\(03\)00336-7](http://doi.org/10.1016/S1053-8119(03)00336-7)
- Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. *Proceedings of the 34th International Conference on Machine Learning*. <https://doi.org/10.48550/arXiv.1701.07875>
- Astrand, E., Enel, P., Ibos, G., Dominey, P. F., Baraduc, P., & Hamed, S. B. (2014). Comparison of Classifiers for Decoding Sensory and Cognitive Information from Prefrontal Neuronal Populations. *PLOS ONE*, 9(1), e86314. <https://doi.org/10.1371/journal.pone.0086314>
- Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41. <https://doi.org/10.1016/j.media.2007.06.004>
- Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., & Hyde, J. S. (1992). Time course EPI of human brain function during task activation. *Mag-*

- netic Resonance in Medicine*, 25(2), 390–397. <https://doi.org/10.1002/mrm.1910250220>
- Bazeille, T. (2021). *Template estimation for arbitrary alignments: Application to brain imaging* [PhD Thesis]. Université Paris-Saclay. <https://www.theses.fr/2021UPASG071>
- Bazeille, T., DuPre, E., Richard, H., Poline, J.-B., & Thirion, B. (2021). An empirical evaluation of functional alignment using inter-subject decoding. *NeuroImage*, 245, 118683. <https://doi.org/10.1016/j.neuroimage.2021.118683>
- Bazeille, T., Richard, H., Janati, H., & Thirion, B. (2019). Local Optimal Transport for Functional Brain Template Estimation. *IPMI 2019 - 26th International Conference on Information Processing in Medical Imaging*. https://doi.org/10.1007/978-3-030-20351-1_18
- Behzadi, Y., Restom, K., Liao, J., & Liu, T. T. (2007). A component based noise correction method (compcor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1), 90–101. <https://doi.org/10.1016/j.neuroimage.2007.04.042>
- Benchetrit, Y., Banville, H., & King, J.-R. (2023). Brain decoding: Toward real-time reconstruction of visual perception. *arXiv*.
- Bhattasali, S., Fabre, M., Luh, W.-M., Al Saied, H., Constant, M., Pallier, C., Brennan, J. R., Spreng, R. N., & Hale, J. (2019). Localising memory retrieval and syntactic composition: An fmri study of naturalistic language comprehension. *Language, Cognition and Neuroscience*, 34(4), 491–510.
- Binda, P., Kurzawski, J. W., Lunghi, C., Biagi, L., Tosetti, M., & Morrone, M. C. (2018). Response to short-term deprivation of the human adult visual cortex measured with 7T BOLD. *Elife*, 7, e40014. <https://elifesciences.org/articles/40014>
- Bouillaud, J. (1825). *Traité clinique et physiologique de l'encéphalite ou Inflammation du cerveau et de ses suites*. <https://gallica.bnf.fr/ark:/12148/bpt6k65257050>
- Bounmy, T., Eger, E., & Meyniel, F. (2023). A characterization of the neural representation of confidence during probabilistic learning. *NeuroImage*, 268, 119849. <https://doi.org/10.1016/j.neuroimage.2022.119849>
- Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4), 375–417. <https://doi.org/10.1002/cpa.3160440402>
- Broca, P. (1865). Perte de la parole, ramouissement chronique et destruction partielle du lobe antérieur gauche du cerveau. <https://www.semanticscholar.org/paper/Perte-de-la-parole%2C-ramouissement-chronique-et-du-Broca/ca2657eabb93b8f6f0011f93525582c7b4d4bb655>
- Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., ... Dale, A. M. (2018). The Adolescent

- Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, 32, 43–54. <https://doi.org/10.1016/j.dcn.2018.03.001>
- Castaldi, E., Lungchi, C., & Morrone, M. C. (2020). Neuroplasticity in adult human visual cortex. *Neuroscience & Biobehavioral Reviews*, 112, 542–552. <https://doi.org/10.1016/j.neubiorev.2020.02.028>
- Chapel, L., Flamary, R., Wu, H., Févotte, C., & Gasso, G. (2021). Unbalanced Optimal Transport through Non-negative Penalized Linear Regression. *Advances in Neural Information Processing Systems*, 34. <http://arxiv.org/abs/2106.04145>
- Chen, P.-H., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., & Ramadge, P. J. (2015). A Reduced-Dimension fMRI Shared Response Model. *Advances in Neural Information Processing Systems*, 28. https://papers.nips.cc/paper_files/paper/2015/hash/b3967a0e938dc2a6340e258630febd5a-Abstract.html
- Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, June). A Simple Framework for Contrastive Learning of Visual Representations. <https://doi.org/10.48550/arXiv.2002.05709>
- Chen, Z., Qing, J., Xiang, T., Yue, W. L., & Zhou, J. H. (2023, March). Seeing Beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding. <https://doi.org/10.48550/arXiv.2211.06956>
- Chen, Z., Qing, J., & Zhou, J. H. (2023, May). Cinematic Mindscapes: High-quality Video Reconstruction from Brain Activity. <https://doi.org/10.48550/arXiv.2305.11675>
- Cheng, F. L., Horikawa, T., Majima, K., Tanaka, M., Abdelhack, M., Aoki, S. C., Hirano, J., & Kamitani, Y. (2023). Reconstructing visual illusory experiences from human brain activity. *Science Advances*, 9(46), eadj3906. <https://doi.org/10.1126/sciadv.adj3906>
- Chizat, L., Peyré, G., Schmitzer, B., & Vialard, F.-X. (2019). Unbalanced Optimal Transport: Dynamic and Kantorovich Formulation. *Journal of Functional Analysis*. <http://arxiv.org/abs/1508.05216>
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123(2), 291–307. <https://academic.oup.com/brain/article-abstract/123/2/291/346042>
- Courty, N., Flamary, R., Tuia, D., & Rakotomamonjy, A. (2016). Optimal Transport for Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <http://arxiv.org/abs/1507.00504>
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, 29(3), 162–173. <https://doi.org/10.1006/cbmr.1996.0014>

- Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26.
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, 9(2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>
- Dantzig, G. B. (1951). Application of the simplex method to a transportation problem. *Activity Analysis & Production & Allocation*. <http://www.ams.org/mathscinet-getitem?mr=56262>
- De Martino, F., Yacoub, E., Kemper, V., Moerel, M., Uludağ, K., De Weerd, P., Ugurbil, K., Goebel, R., & Formisano, E. (2018). The impact of ultra-high field MRI on cognitive and computational neuroimaging. *NeuroImage*, 168, 366–382. <https://doi.org/10.1016/j.neuroimage.2017.03.060>
- Défossez, A., Caucheteux, C., Rapin, J., Kabeli, O., & King, J.-R. (2023). Decoding speech from non-invasive brain recordings. <https://doi.org/10.48550/arXiv.2208.12266>
- Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences*, 9(7), 335–341. <https://doi.org/10.1016/j.tics.2005.05.004>
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1), 1–15. <https://doi.org/10.1016/j.neuroimage.2010.06.010>
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science*, 293(5539), 2470–2473. <https://doi.org/10.1126/science.1063414>
- Dumoulin, S. O., Fracasso, A., van der Zwaag, W., Siero, J. C., & Petridou, N. (2018). Ultra-high field MRI: Advancing systems neuroscience towards mesoscopic human brain function. *NeuroImage*, 168, 345–357. <https://doi.org/10.1016/j.neuroimage.2017.01.028>
- Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in neurosciences*, 23(10), 475–483. [https://www.cell.com/trends/neurosciences/fulltext/S0166-2236\(00\)01633-7](https://www.cell.com/trends/neurosciences/fulltext/S0166-2236(00)01633-7)
- Eichert, N., Robinson, E. C., Bryant, K. L., Jbabdi, S., Jenkinson, M., Li, L., Krug, K., Watkins, K. E., & Mars, R. B. (2020). Cross-species cortical alignment identifies different types of anatomical reorganization in the primate temporal lobe. *eLife*, 9, e53232. <https://doi.org/10.7554/eLife.53232>
- Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, 23(1), 115–125. [https://www.cell.com/neuron/pdf/S0896-6273\(00\)80758-8.pdf](https://www.cell.com/neuron/pdf/S0896-6273(00)80758-8.pdf)
- Erb, J., Armendariz, M., De Martino, F., Goebel, R., Vanduffel, W., & Formisano, E. (2019). Homology and specificity of natural sound-encoding in human and monkey auditory cortex. *Cerebral cortex*, 29(9), 3636–3650.

- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. <https://doi.org/10.1038/s41592-018-0235-4>
- Fedorenko, E. (2021). The early origins and the growing popularity of the individual-subject analytic approach in human neuroscience. *Current Opinion in Behavioral Sciences*, *40*, 105–112. <https://doi.org/10.1016/j.cobeha.2021.02.023>
- Feilong, M., Nastase, S. A., Jiahui, G., Halchenko, Y. O., Gobbini, M. I., & Haxby, J. V. (2022, May). The Individualized Neural Tuning Model: Precise and generalizable cartography of functional architecture in individual brains. <https://doi.org/10.1101/2022.05.15.492022>
- Ferrante, M., Boccato, T., & Toschi, N. (2023, August). Through their eyes: Multi-subject Brain Decoding with simple alignment techniques. <https://doi.org/10.48550/arXiv.2309.00627>
- Ferrante, M., Ozcelik, F., Boccato, T., VanRullen, R., & Toschi, N. (2023, May). Brain Captioning: Decoding human brain activity into images and text. <https://doi.org/10.48550/arXiv.2305.11560>
- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. <https://doi.org/10.1038/nn.4135>
- Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>
- Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp*, *8*(4), 272–284. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:4<3C272::AID-HBM10%3E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<3C272::AID-HBM10%3E3.0.CO;2-4)
- Flourens, P. (1824). *Recherches expérimentales sur les propriétés et les fonctions du système nerveux dans les animaux vertébrés*. Crevot. <http://archive.org/details/rcherchesexprim00flougoog>
- Forrow, A., Hütter, J.-C., Nitzan, M., Rigollet, P., Schiebinger, G., & Weed, J. (2018). Statistical Optimal Transport via Factored Couplings. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. <http://arxiv.org/abs/1806.07348>
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*(4), 189–210. <https://doi.org/10.1002/hbm.460020402>
- Friston, K., Frith, C., Frackowiak, R., & Turner, R. (1995). Characterizing Dynamic Brain Responses with fMRI: A Multivariate Approach. *Neuroimage*, *2*(2), 166–172. <https://doi.org/10.1006/nimg.1995.1019>

- Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. <http://doi.org/10.1038/nature18933>
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., Jenkinson, M., & WU-Minn HCP Consortium. (2013). The minimal pre-processing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. <https://doi.org/10.1016/j.neuroimage.2013.04.127>
- Glickstein, M., & Fahle, M. (2000). T. Inouye: Visual disturbances following gunshot wounds of the cortical visual area (Translation). *Brain*, *123*.
- Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*(4), 416–429. <https://doi.org/10.1006/nimg.1998.0419>
- Gramfort, A., Peyré, G., & Cuturi, M. (2015). Fast optimal transport averaging of neuroimaging data. <https://doi.org/10.48550/ARXIV.1503.08596>
- Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., Dosenbach, N. U. F., & Petersen, S. E. (2018). Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*, *98*(2), 439–452.e5. <https://doi.org/10.1016/j.neuron.2018.03.035>
- Gu, Z., Jamison, K., Kuceyeski, A., & Sabuncu, M. (2023, March). Decoding natural image stimuli from fMRI data with a surface-based convolutional network. <https://doi.org/10.48550/arXiv.2212.02409>
- Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex [Publisher: Oxford Academic]. *Cerebral Cortex*, *26*(6), 2919–2934. <https://doi.org/10.1093/cercor/bhw068>
- Hamed, S. B., Schieber, M. H., & Pouget, A. (2007). Decoding M1 Neurons During Multiple Finger Movements. *Journal of Neurophysiology*, *98*(1), 327–333. <https://doi.org/10.1152/jn.00760.2006>
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)*, *293*(5539), 2425–2430. <https://doi.org/10.1126/science.1063736>
- Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416. <https://doi.org/10.1016/j.neuron.2011.08.026>

- Haxby, J. V., Guntupalli, J. S., Nastase, S. A., & Feilong, M. (2020). Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. *eLife*, 9, e56601. <https://doi.org/10.7554/eLife.56601>
- Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature reviews neuroscience*, 7(7), 523–534.
- Hitchcock, F. L. (1941). The Distribution of a Product from Several Sources to Numerous Localities. *Journal of Mathematics and Physics*, 20(1-4), 224–230. <https://doi.org/10.1002/sapm1941201224>
- Ho, J. K., Horikawa, T., Majima, K., Cheng, F., & Kamitani, Y. (2023). Inter-individual deep image reconstruction via hierarchical neural code conversion. *NeuroImage*, 271, 120007. <https://doi.org/10.1016/j.neuroimage.2023.120007>
- Holmes, G. (1917). The symptoms of acute cerebellar injuries due to gunshot injuries. *Brain: A Journal of Neurology*, 40, 461–535. <https://doi.org/10.1093/brain/40.4.461>
- Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural Decoding of Visual Imagery During Sleep. *Science*, 340(6132), 639–642. <https://doi.org/10.1126/science.1234330>
- Horikawa, T., & Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features. *Nat. Commun.*, 8(15037), 1–15. <https://doi.org/10.1038/ncomms15037>
- Huang, R.-S., & Sereno, M. I. (2007). Dodecapus: An MR-compatible system for somatosensory stimulation. *NeuroImage*, 34(3), 1060–1073. <https://doi.org/10.1016/j.neuroimage.2006.10.024>
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453–458. <https://doi.org/10.1038/nature17637>
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143–156. [https://doi.org/10.1016/s1361-8415\(01\)00036-6](https://doi.org/10.1016/s1361-8415(01)00036-6)
- Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., & Orban, G. A. (2012). Processing of vocalizations in humans and monkeys: A comparative fMRI study. *Neuroimage*, 62(3), 1376–1389.
- Kantorovich, L. V. (1942). On the translocation of masses. *Dokl. Akad. Nauk SSSR*, 37(7-8), 199–201. <https://cds.cern.ch/record/739801>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, 17(11), 4302–4311. <https://www.jneurosci.org/content/17/11/4302.short>
- Kolster, H., Mandeville, J. B., Arsenault, J. T., Ekstrom, L. B., Wald, L. L., & Vanduffel, W. (2009). Visual Field Map Clusters in Macaque Extrastriate Visual Cortex [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, 29(21), 7031–7039. <https://doi.org/10.1523/JNEUROSCI.0518-09.2009>

- Kolster, H., Peeters, R., & Orban, G. A. (2010). The Retinotopic Organization of the Human Middle Temporal Area MT/V5 and Its Cortical Neighbors. *Journal of Neuroscience*, *30*(29), 9801–9820.
- Komodakis, N., & Tziritas, G. (2007). Approximate Labeling via Graph Cuts Based on Linear Programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(8), 1436–1453. <https://doi.org/10.1109/TPAMI.2007.1061>
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. <https://doi.org/10.1016/j.neuron.2008.10.043>
- Kupersmidt, G., Belyi, R., Gaziv, G., & Irani, M. (2022, June). A Penny for Your (visual) Thoughts: Self-Supervised Reconstruction of Natural Movies from Brain Activity. <https://doi.org/10.48550/arXiv.2206.03544>
- Lahner, B., Dwivedi, K., Iamshchinina, P., Graumann, M., Lascelles, A., Roig, G., Gifford, A. T., Pan, B., Jin, S., Murty, N. A. R., Kay, K., Oliva, A., & Cichy, R. (2023, March). BOLD Moments: Modeling short visual events through a video fMRI dataset and metadata. <https://doi.org/10.1101/2023.03.12.530887>
- LeBel, A., Wagner, L., Jain, S., Adhikari-Desai, A., Gupta, B., Morgenthal, A., Tang, J., Xu, L., & Huth, A. G. (2023). A natural language fMRI dataset for voxelwise encoding models. *Scientific Data*, *10*(1), 555. <https://doi.org/10.1038/s41597-023-02437-z>
- Leite, F. P., Tsao, D., Vanduffel, W., Fize, D., Sasaki, Y., Wald, L. L., Dale, A. M., Kwong, K. K., Orban, G. A., & Rosen, B. R. (2002). Repeated fMRI using iron oxide contrast agent in awake, behaving macaques at 3 Tesla. *Neuroimage*, *16*(2), 283–294.
- Liero, M., Mielke, A., & Savaré, G. (2018). Optimal Entropy-Transport problems and a new Hellinger–Kantorovich distance between positive measures. *Inventiones mathematicae*, *211*(3), 969–1117. <https://doi.org/10.1007/s00222-017-0759-8>
- Madan, C. R. (2022). Scan Once, Analyse Many: Using Large Open-Access Neuroimaging Datasets to Understand the Brain. *Neuroinformatics*, *20*(1), 109–137. <https://doi.org/10.1007/s12021-021-09519-6>
- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., & Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, *16*(2), 187–198. <https://doi.org/10.1109/42.563664>
- Mai, W., & Zhang, Z. (2023, August). UniBrain: Unify Image Reconstruction and Captioning All in One Diffusion Model from Human Brain Activity. <https://doi.org/10.48550/arXiv.2308.07428>
- Mantini, D., Hasson, U., Betti, V., Perrucci, M. G., Romani, G. L., Corbetta, M., Orban, G. A., & Vanduffel, W. (2012). Interspecies activity correlations reveal functional correspondence between monkey and human brain ar-

- ... eas. *Nature Methods*, 9(3), 277–282. <https://doi.org/10.1038/nmeth.1868>
- Mars, R. B., Sotiropoulos, S. N., Passingham, R. E., Sallet, J., Verhagen, L., Khrapitchev, A. A., Sibson, N., & Jbabdi, S. (2018). Whole brain comparative anatomy using connectivity blueprints. *eLife*, 7, e35237. <https://doi.org/10.7554/eLife.35237>
- Mars, R. B., Verhagen, L., Gladwin, T. E., Neubert, F.-X., Sallet, J., & Rushworth, M. F. S. (2016). Comparing brains by matching connectivity profiles. *Neuroscience and Biobehavioral Reviews*, 60, 90–97. <https://doi.org/10.1016/j.neubiorev.2015.10.008>
- Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., & Lancaster, J. (1995). A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development: The International Consortium for Brain Mapping (ICBM). *NeuroImage*, 2(2, Part A), 89–101. <https://doi.org/10.1006/nimg.1995.1012>
- Memoli, F. (2007). *On the use of Gromov-Hausdorff Distances for Shape Comparison*. The Eurographics Association. <https://doi.org/10.2312/SPBG/SPBG07/081-090>
- Mémoli, F. (2011). Gromov–Wasserstein Distances and the Metric Approach to Object Matching. *Foundations of Computational Mathematics*, 11(4), 417–487. <https://doi.org/10.1007/s10208-011-9093-5>
- Mensch, A., Mairal, J., Thirion, B., & Varoquaux, G. (2021). Extracting representations of cognition across neuroimaging studies improves brain decoding. *PLoS computational biology*, 17(5), e1008795. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008795>
- Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., & Newman, S. (2004). Learning to decode cognitive states from brain images. *Machine learning*, 57, 145–175.
- Moerel, M., De Martino, F., & Formisano, E. (2012). Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning, and relation to voice sensitivity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(41), 14205–14216. <https://doi.org/10.1523/JNEUROSCI.1388-12.2012>
- Mokady, R., Hertz, A., & Bermano, A. H. (2021, November). ClipCap: CLIP Prefix for Image Captioning. <https://doi.org/10.48550/arXiv.2111.09734>
- Monge, G. (1781). *Mémoire sur la théorie des déblais et des remblais*. Imprimerie royale.
- Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, 222, 117254. <https://doi.org/10.1016/j.neuroimage.2020.117254>
- Neubert, F.-X., Mars, R. B., Thomas, A. G., Sallet, J., & Rushworth, M. F. S. (2014). Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron*, 81(3), 700–713. <https://doi.org/10.1016/j.neuron.2013.11.012>

- Nishimoto, S., Vu, A., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. (2011). Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. *Curr Biol*, 21, 1641–6. <https://doi.org/10.1016/j.cub.2011.08.031>
- Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). 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. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC55275/>
- Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, 89(13), 5951–5955. <https://doi.org/10.1073/pnas.89.13.5951>
- Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences*, 8(7), 315–324. <https://doi.org/10.1016/j.tics.2004.05.009>
- Ozcelik, F., & VanRullen, R. (2023, June). Natural scene reconstruction from fMRI signals using generative latent diffusion. <https://doi.org/10.48550/arXiv.2303.05334>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox & R. Garnett (Eds.), *Advances in neural information processing systems* 32 (pp. 8024–8035). <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
- Pele, O., & Werman, M. (2009). Fast and Robust Earth Mover's Distances, 460–467. <https://doi.org/10.1109/ICCV.2009.5459199>
- Peyré, G., & Cuturi, M. (2020). Computational Optimal Transport. <http://arxiv.org/abs/1803.00567>
- Peyré, G., Cuturi, M., & Solomon, J. (2016). Gromov-Wasserstein Averaging of Kernel and Distance Matrices. *Proceedings of The 33rd International Conference on Machine Learning*, 2664–2672. <https://proceedings.mlr.press/v48/peyre16.html>
- Phillips, E. M., Gillette, K. D., Dilks, D. D., & Berns, G. S. (2022). Through a Dog's Eyes: fMRI Decoding of Naturalistic Videos from the Dog Cortex. *JoVE*

- (*Journal of Visualized Experiments*), (187), e64442. <https://doi.org/10.3791/64442>
- Pinho, A. L., Amadon, A., Fabre, M., Dohmatob, E., Denghien, I., Torre, J. J., Ginisty, C., Becuwe-Desmidt, S., Roger, S., Laurier, L., Joly-Testault, V., Médiouni-Cloarec, G., Doublé, C., Martins, B., Pinel, P., Eger, E., Varoquaux, G., Pallier, C., Dehaene, S., ... Thirion, B. (2021). Subject-specific segregation of functional territories based on deep phenotyping. *Hum Brain Mapp*, 42(4), 841–870. <https://doi.org/10.1002/hbm.25189>
- Pinho, A. L., Amadon, A., Ruest, T., Fabre, M., Dohmatob, E., Denghien, I., Ginisty, C., Becuwe-Desmidt, S., Roger, S., Laurier, L., Joly-Testault, V., Médiouni-Cloarec, G., Doublé, C., Martins, B., Pinel, P., Eger, E., Varoquaux, G., Pallier, C., Dehaene, S., ... Thirion, B. (2018). Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. *Scientific Data*, 5(1), 180105. <https://doi.org/10.1038/sdata.2018.105>
- Pinho, A. L., Richard, H., Eickenberg, M., Amadon, A., Dohmatob, E., Denghien, I., Torre, J. J., Shankar, S., Aggarwal, H., & Ponce, A. F. (2023). Individual Brain Charting third release, probing brain activity during Movie Watching and Retinotopic Mapping. <https://hal.science/hal-04272993/>
- Pizzagalli, F., Auzias, G., Yang, Q., Mathias, S. R., Faskowitz, J., Boyd, J. D., Amini, A., Rivière, D., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., Mangin, J. F., Glahn, D. C., Blangero, J., Wright, M. J., Thompson, P. M., Kochunov, P., & Jahanshad, N. (2020). The reliability and heritability of cortical folds and their genetic correlations across hemispheres. *Commun Biol*, 3(1), 510.
- Platt, J. (2005). FastMap, MetricMap, and Landmark MDS are all Nystrom Algorithms. <https://www.microsoft.com/en-us/research/publication/fastmap-metricmap-and-landmark-mds-are-all-nystrom-algorithms/>
- Popivanov, I. D., Jastorff, J., Vanduffel, W., & Vogels, R. (2012). Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. *Neuroimage*, 63(2), 723–741.
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021, February). Learning Transferable Visual Models From Natural Language Supervision. <https://doi.org/10.48550/arXiv.2103.00020>
- Richard, H., Gresele, L., Hyvarinen, A., Thirion, B., Gramfort, A., & Ablin, P. (2020). Modeling shared responses in neuroimaging studies through multiview ica. *Advances in Neural Information Processing Systems*, 33, 19149–19162.
- Richard, H., Gresele, L., Hyvärinen, A., Thirion, B., Gramfort, A., & Ablin, P. (2020, December). Modeling Shared Responses in Neuroimaging Studies through MultiView ICA. <https://doi.org/10.48550/arXiv.2006.06635>
- Robinson, E. C., Garcia, K., Glasser, M. F., Chen, Z., Coalson, T. S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., Hutter, J., Price, A., Cordero Grande, L., Hughes, E., Tumor, N., Bayly, P. V., Van Essen, D. C.,

- Smith, S. M., Edwards, A. D., . . . Rueckert, D. (2018). Multimodal surface matching with higher-order smoothness constraints. *NeuroImage*, *167*, 453–465. <https://doi.org/10.1016/j.neuroimage.2017.10.037>
- Robinson, E. C., Jbabdi, S., Glasser, M. F., Andersson, J., Burgess, G. C., Harms, M. P., Smith, S. M., Van Essen, D. C., & Jenkinson, M. (2014). MSM: A new flexible framework for Multimodal Surface Matching. *NeuroImage*, *100*, 414–426. <https://doi.org/10.1016/j.neuroimage.2014.05.069>
- Sabuncu, M. R., Singer, B. D., Conroy, B., Bryan, R. E., Ramadge, P. J., & Haxby, J. V. (2010). Function-based intersubject alignment of human cortical anatomy. *Cerebral Cortex (New York, N.Y.: 1991)*, *20*(1), 130–140. <https://doi.org/10.1093/cercor/bhp085>
- Saxe, R., & Kanwisher, N. (2013). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind”. In *Social neuroscience* (pp. 171–182). Psychology Press. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203496190-20/people-thinking-thinking-people-saxe-kanwisher>
- Scetbon, M., Cuturi, M., & Peyré, G. (2021, March). Low-Rank Sinkhorn Factorization. <https://doi.org/10.48550/arXiv.2103.04737>
- Scetbon, M., Klein, M., Palla, G., & Cuturi, M. (2023, May). Unbalanced Low-rank Optimal Transport Solvers. <https://doi.org/10.48550/arXiv.2305.19727>
- Scetbon, M., Peyré, G., & Cuturi, M. (2021). Linear-Time Gromov Wasserstein Distances using Low Rank Couplings and Costs. <http://arxiv.org/abs/2106.01128>
- Schalk, G., Kapeller, C., Guger, C., Ogawa, H., Hiroshima, S., Lafer-Sousa, R., Saygin, Z. M., Kamada, K., & Kanwisher, N. (2017). Facephenes and rainbows: Causal evidence for functional and anatomical specificity of face and color processing in the human brain. *Proceedings of the National Academy of Sciences*, *114*(46), 12285–12290. <https://doi.org/10.1073/pnas.1713447114>
- Schneider, M., Kemper, V. G., Emmerling, T. C., De Martino, F., & Goebel, R. (2019). Columnar clusters in the human motion complex reflect consciously perceived motion axis. *Proceedings of the National Academy of Sciences*, *116*(11), 5096–5101. <https://doi.org/10.1073/pnas.1814504116>
- Scotti, P. S., Banerjee, A., Goode, J., Shabalin, S., Nguyen, A., Cohen, E., Dempster, A. J., Verlinde, N., Yundler, E., Weisberg, D., Norman, K. A., & Abraham, T. M. (2023, May). Reconstructing the Mind’s Eye: fMRI-to-Image with Contrastive Learning and Diffusion Priors. <http://arxiv.org/abs/2305.18274>
- Scotti, P. S., Tripathy, M., Villanueva, C. K. T., Kneeland, R., Chen, T., Narang, A., Santhirasegaran, C., Xu, J., Naselaris, T., Norman, K. A., & Abraham, T. M. (2024, March). MindEye2: Shared-Subject Models Enable fMRI-To-Image With 1 Hour of Data. <http://arxiv.org/abs/2403.11207>

- Séjourné, T., Vialard, F.-X., & Peyré, G. (2021). The Unbalanced Gromov Wasserstein Distance: Conic Formulation and Relaxation. *Advances in Neural Information Processing Systems*, 34. <http://arxiv.org/abs/2009.04266>
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science (New York, N.Y.)*, 268(5212), 889–893. <https://doi.org/10.1126/science.7754376>
- Sliwa, J., & Freiwald, W. (2017). A dedicated network for social interaction processing in the primate brain. *Science*, 356, 745–749. <https://doi.org/10.1126/science.aam6383>
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3), e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
- Takagi, Y., & Nishimoto, S. (2023, March). High-resolution image reconstruction with latent diffusion models from human brain activity. <https://doi.org/10.1101/2022.11.18.517004>
- Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings [Number: 5 Publisher: Nature Publishing Group]. *Nature Neuroscience*, 26(5), 858–866. <https://doi.org/10.1038/s41593-023-01304-9>
- Tavor, I., Jones, O. P., Mars, R. B., Smith, S., Behrens, T., & Jbabdi, S. (2016). Task-free mri predicts individual differences in brain activity during task performance. *Science*, 352(6282), 216–220.
- Thirion, B., Aggarwal, H., Ponce, A. F., Pinho, A. L., & Thual, A. (2023). Should one go for individual- or group-level brain parcellations? A deep-phenotyping benchmark. *Brain Structure and Function*, 229(1), 161–181. <https://doi.org/10.1007/s00429-023-02723-x>
- Thirion, B., Thual, A., & Pinho, A. L. (2021). From deep brain phenotyping to functional atlas. *Current Opinion in Behavioral Sciences*, 40, 201–212. <https://doi.org/10.1016/j.cobeha.2021.05.004>
- Thirion, B., Varoquaux, G., Dohmatob, E., & Poline, J.-B. (2014). Which fMRI clustering gives good brain parcellations? *Frontiers in Neuroscience*, 8(167), 13. <https://doi.org/10.3389/fnins.2014.00167>
- Thirion, J.-P. (1998). Image matching as a diffusion process: An analogy with Maxwell's demons. *Medical Image Analysis*, 2(3), 243–260. [https://doi.org/10.1016/S1361-8415\(98\)80022-4](https://doi.org/10.1016/S1361-8415(98)80022-4)
- Thomas, A., Ré, C., & Poldrack, R. (2022). Self-Supervised Learning of Brain Dynamics from Broad Neuroimaging Data. *Advances in Neural Information Processing Systems*, 35, 21255–21269. <https://doi.org/10.26434/chemrxiv-2022-08>

- [proceedings . neurips . cc / paper _ files / paper / 2022 / hash / 8600a9df1a087a9a66900cc8c948c3f0-Abstract-Conference.html](https://proceedings.neurips.cc/paper_files/paper/2022/hash/8600a9df1a087a9a66900cc8c948c3f0-Abstract-Conference.html)
- Thual, A., Benchetrit, Y., Geilert, F., Rapin, J., Makarov, I., Banville, H., & King, J.-R. (2023). Aligning brain functions boosts the decoding of visual semantics in novel subjects. <https://doi.org/10.48550/arXiv.2312.06467>
- Thual, A., Tran, Q. H., Zemskova, T., Courty, N., Flamary, R., Dehaene, S., & Thirion, B. (2022). Aligning individual brains with fused unbalanced Gromov Wasserstein. *Advances in Neural Information Processing Systems*, 35, 21792–21804. https://proceedings.neurips.cc/paper_files/paper/2022/hash/8906cac4ca58dcdf17e97a0486ad57ca-Abstract-Conference.html
- Tolstoi, A. (1930). Metody nakhozheniya naimen'shego summovogo kilometrazha pri planirovanii perevozok v prostranstve [Russian; Methods of finding the minimal total kilometrage in cargo- transportation planning in space. *Planirovanie Perevozok, Sbornik pervy'ı* [Russian; Transportation Planning, Volume I.
- Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, 6(9), 989–995. <https://doi.org/10.1038/nn1111>
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A Cortical Region Consisting Entirely of Face-Selective Cells. *Science (New York, N.Y.)*, 311(5761), 670–674. <https://doi.org/10.1126/science.1119983>
- Tsao, D. Y., Schweers, N., Moeller, S., & Freiwald, W. A. (2008). Patches of face-selective cortex in the macaque frontal lobe. *Nature Neuroscience*, 11(8), 877–879. <https://doi.org/10.1038/nn.2158>
- Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., & Coalson, T. (2012). Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cereb Cortex*, 22(10), 2241–2262.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Van Essen, D., Barch, D., Corbetta, M., Goate, A., Heath, A., Larson-Prior, L., Marcus, D., Petersen, S., Prior, F., Province, M., Raichle, M., Schlaggar, B., Shimony, J., ... Salmelin, R. (2013). The WU-Minn Human Connectome Project: an overview. *Neuroimage*, 80, 62–79.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An Overview. *NeuroImage*, 80, 62–79. <https://doi.org/10.1016/j.neuroimage.2013.05.041>
- Vanduffel, W., Fize, D., Mandeville, J. B., Nelissen, K., Van Hecke, P., Rosen, B. R., Tootell, R. B., & Orban, G. A. (2001). Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron*, 32(4), 565–577.
- Vayer, T., Chapel, L., Flamary, R., Tavenard, R., & Courty, N. (2020). Fused Gromov-Wasserstein Distance for Structured Objects. *Algorithms*, 13(9), 212. <https://doi.org/10.3390/a13090212>

- Villani, C. (2003). *Topics in Optimal Transportation*. American Mathematical Soc.
- Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-Form System. *Neuron*, 55(1), 143–156. <https://doi.org/10.1016/j.neuron.2007.05.031>
- Vizioli, L., Moeller, S., Dowdle, L., Akçakaya, M., De Martino, F., Yacoub, E., & Uğurbil, K. (2021). Lowering the thermal noise barrier in functional brain mapping with magnetic resonance imaging. *Nature Communications*, 12(1), 5181. <https://doi.org/10.1038/s41467-021-25431-8>
- Wang, C., Yan, H., Huang, W., Li, J., Wang, Y., Fan, Y.-S., Sheng, W., Liu, T., Li, R., & Chen, H. (2022). Reconstructing rapid natural vision with fMRI-conditional video generative adversarial network. *Cerebral Cortex*, 32(20), 4502–4511. <https://doi.org/10.1093/cercor/bhab498>
- Wen, H., Shi, J., Zhang, Y., Lu, K.-H., Cao, J., & Liu, Z. (2017, September). Data for Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision Tests. <https://doi.org/doi:10.4231/R7SF2TCW>
- Wen, H., Shi, J., Zhang, Y., Lu, K.-H., Cao, J., & Liu, Z. (2018). Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision. *Cerebral Cortex (New York, N.Y.: 1991)*, 28(12), 4136–4160. <https://doi.org/10.1093/cercor/bhx268>
- Wilson, A. G. (1969). The Use of Entropy Maximising Models, in the Theory of Trip Distribution, Mode Split and Route Split. *Journal of Transport Economics and Policy*, 3(1), 108–126. <https://www.jstor.org/stable/20052128>
- Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, 14(6), 1370–1386. <https://doi.org/10.1006/nimg.2001.0931>
- Xie, Y., Wang, X., Wang, R., & Zha, H. (2020). A Fast Proximal Point Method for Computing Exact Wasserstein Distance. *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, 433–453. <https://proceedings.mlr.press/v115/xie20b.html>
- Xu, T., Nenning, K.-H., Schwartz, E., Hong, S.-J., Vogelstein, J. T., Goulas, A., Fair, D. A., Schroeder, C. E., Margulies, D. S., Smallwood, J., Milham, M. P., & Langs, G. (2020). Cross-species functional alignment reveals evolutionary hierarchy within the connectome. *NeuroImage*, 223, 117346. <https://doi.org/10.1016/j.neuroimage.2020.117346>
- Yeo, B. T., Sabuncu, M. R., Vercauteren, T., Ayache, N., Fischl, B., & Golland, P. (2010). Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration. *IEEE transactions on medical imaging*, 29(3), 650–668. <https://doi.org/10.1109/TMI.2009.2030797>
- Zhan, M., Pallier, C., Agrawal, A., Dehaene, S., & Cohen, L. (2023). Does the visual word form area split in bilingual readers? A millimeter-scale 7-T fMRI

study. *Science Advances*, 9(14), eadf6140. <https://doi.org/10.1126/sciadv.adf6140>

Zhang, T., Gao, J. S., Cukur, T., & Gallant, J. L. (2020). Whole-brain BOLD activity recorded by fMRI during a visual attention task and a video game task. <https://doi.org/10.6080/K0668BDF>