Towards a learning fingerprint: new methods and paradigms for complex motor skill learning in fMRI

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von
Eric Lacosse
Chicago, U.S.A.

November, 2020

Tag der mündlichen Prüfung: 23. Februar 2021

Stellv. Dekan der Math.-Nat. Fakultät: Prof. Dr. József Fortágh
Dekan der Medizinischen Fakultät: Prof. Dr. Bernd Pichler

1. Berichterstatter: Klaus Scheffler
2. Berichterstatter: Marc Himmelbach

Prüfungskommission:
Prof. Dr. Klaus Scheffler
PD Dr. Marc Himmelbach
PD Dr. Gabriele Lohmann
Dr. Georg Martius

Declaration

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel: „Towards a learning fingerprint: new methods and paradigms for complex motor skill learning in fMRI" selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

I, Eric Lacosse, hereby declare that I have produced the work entitled:
"Towards a learning fingerprint: new methods and paradigms for complex motor skill learning in fMRI"
Submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, den ....
Datum/Date   *Unterschrift/Signature*

Abstract

Functional Magnetic Resonance Imaging (fMRI) research in sensorimotor learning focus on two separate paradigms: (1) task-based (tfMRI), where brain changes are evaluated according to activity elicited by performance of the task, or (2) task-free, i.e., resting-state (rsfMRI), where changes are reflected in spontaneous, internally generated brain activity. While the former paradigm allows careful control and manipulation of the task, the later allows unrestrained motor learning tasks to take place beyond the limitations of the scanner environment. Machine learning approaches attempting to model these two types of measurements together to explain physiological effects of learning remained unexplored. Although these paradigms yield results showing considerable overlap between their topographical patterns, they are usually treated separately. Consequently, their relationship, and how or if any behaviorally relevant neural information processing mediates it, remains unclear. To resolve this ambiguity, new methodology was developed guided by questions of sensorimotor learning in motor tasks having dynamics completely specified mathematically.

First, basic fMRI methodological considerations were made. Machine learning methods that claimed to predict individual tfMRI task maps from rsfMRI activity were improved. In reviewing previous methodology, most methods were found to underperform against trivial baseline model performances based on massive group averaging. New methods were developed that remedies this problem to a great extent. Benchmark comparisons and model evaluation metrics demonstrating empirical properties related to this predictive mapping previously unconsidered were also further developed. With these newly formed empirical observations, a relationship between individual prediction scores and behavioral performance measured during the task could be established.

Second, a complex motor learning task performed during an fMRI measurement was designed to relate learning effects observed in both types of measurements from a single longitudinal learning session. Participants measured while performing the task show they learn to exploit a property that drives brain activity in certain regions towards a state requiring less active

control and error correction. Reconfiguration of functional activity in task-evoked and task-free activity from these behavioral learning effects were investigated, applying methodology developed earlier in an attempt to relate them together. Predictions of individual task-evoked responses from rsfMRI provide a relative measure of dependence, however, remain limited for reasons understood from the methodological study. No rsfMRI reconfiguration due to learning was detected, yet changes over the course of learning in task-evoked activity appear significant. Increasing recruitment of the Default Mode Network (DMN) during the task explain these changes. These results support that minimal reconfiguration of the cortex suggestive of plasticity effects are needed to find task solutions in a passively stable space.

Table of Contents

Chapter 1

Introduction

"A potential objection to [the idea that a higher-level conceptualization are necessary, i.e., a description of a behavior] might be to say, 'Who cares what philosophers say about the differences between psychology and neuroscience, or reductionism in general? We are scientists, not philosophers!' The answer to this is simple: there is no escape from philosophy. Every scientist takes a philosophical position, either tacitly or explicitly, whenever they state that a result is 'important,' 'fundamental,' or 'interesting.' This is because such assertions are always a judgment from outside of science. There is no 'interesting' variable inherent to the data that can be objectively plotted on a graph—abstract reasoning and normative claims cannot be substituted by, or obtained from, data."

Krakauer et al., a perspective on why neuroscience needs behavior (Krakauer et al. 2017).

1.1   Taking on Two Problems in the Field

The relationship between the Central Nervous System (CNS) and environment is extremely versatile. This versatility is what underlies all human learning. It is also what allows the CNS to gain so much from so little–so little energy, time, and data (sensory information and experience). This stands in stark contrast with today's most advanced approaches of synthetic intelligence; enormous amounts of compute time, energy, and data are needed for the most rudimentary of effortless, commonsense tasks we take for granted.

One inspiring, general category of tasks–universally celebrated across cultures–is learning how to generate or adapt skilled movement, i.e., *motor learning.* And yet, we are far from understanding how a biological feat like learning a new motor skill takes place–even in the

simplest form conceivable. We also struggle to realize it synthetically–robots performing the simplest tasks still remain far behind human dexterity, reliability, and generalizability. Why is it so difficult? Do current approaches to understand motor learning place its focus in the wrong places?

This thesis is a culmination of my own thought that brings forward a new approach–from the design of improved methodology mapping resting-state and task fMRI measurements together to studying a complex motor task that has not yet been examined with fMRI. Together, the overall approach described in this chapter worked towards correcting two biases I believe to be present in current research to study sensorimotor skill learning.

The first bias is ignoring the role of internally generated activity in shaping, or becoming shaped by, brain activation elicited during a task. A considerable accumulation of literature suggests brain activity not only depends on external stimulus information, but is heavily influenced by internally generated activity (Northoff et al. 2010). Further, how internally generated activity relates to activity evoked externally may fundamentally be shaped by Hebbian-like plasticity processes. Methodology with the specificity to identify those effects remain underdeveloped; what actually constitutes a signal of interest rather than noise in recorded brain activity is a contentious topic (Uddin 2020). This is especially the case for the research focus presented here on fMRI measurement. Therefore, it remains difficult to isolate signals that reflect truly important internally generated activity, especially when simultaneous electrophysiological and fMRI investigations remain sparse. So far, countless fMRI studies have either focused on measuring task-activity or internally generated activity, i.e., resting-state, *independently*. This is especially the case for most motor skill learning investigations. Showing that activation or functional connectivity (FC) patterns significantly differ between task conditions or among groups of individuals in either of these measurements often falls short of implicit goals; characterizations stopping there remain vague and are increasingly recognized to lack a substantial neuroscientific value (Kullmann 2020). Analyses need to be constrained by theories of neural information processing that involve not only *behavior*, but plausible physiological mechanisms (Jonas & Kording 2017) informed by multi-level investigation (Logothetis 2008). The independent foci of the fMRI community on either internally generated activity or task-generated activity leaves a chasm to be bridged. What lacks are studies relying on a more specific methodology showing how patterns of activity engaged in a task reflect patterns of internally generated activity. Furthermore (as the opening quote to this introduction chapter highlights), they need to be grounded in research questions that are consequential to explain behavior. These kinds of fMRI studies remain few and far between.
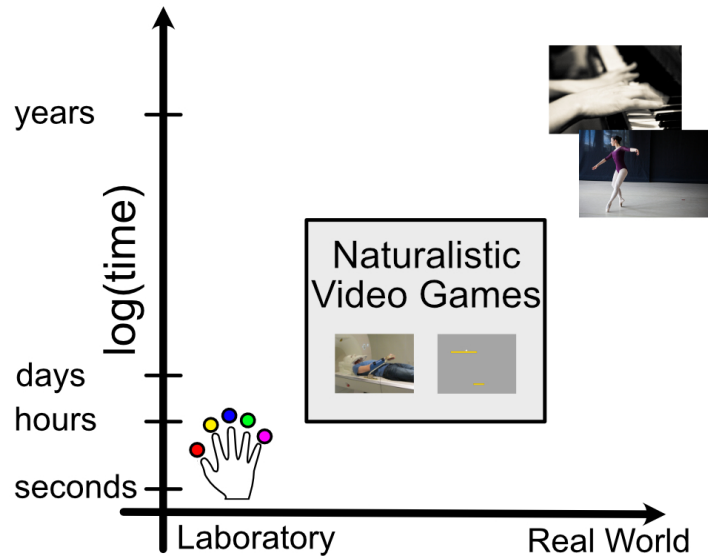
Figure 1.1: Lab tasks versus real-world motor behavior. Most motor skill learning tasks examined in laboratory settings are highly abstracted away from any real-world behavior we wish to explain. Compare sensorimotor adaptation or sequence learning task paradigms to playing the piano or ballet dancing–the earlier does not even remotely approach the complexity or skill required for either of the later. Unfortunately, the motor behavior we wish to be able to explain at the end of the day still cannot be brought into the scanner. Therefore, video games involving ecologically motivated behavior might be a healthy compromise between the two extremes.

The second bias is a hesitation to study complex skills that more closely resemble real-world behavior, figure 1.1. Most experimental paradigms rely on simple tasks, e.g., reaching for a target or generating a finger sequence. This is especially the case within the fMRI community where mobility constraints during measurement impose severe limitations. While these paradigms simplify the design of motor control hypotheses, isolating alleged cognitive processes, they do not necessarily generalize well beyond their highly controlled laboratory settings (Wulf & Shea 2002). This is problematic, especially when the MR scanner environment itself is a highly unnatural setting for examining human behavior. Additionally, simple tasks easy to study and decompose in a lab environment appear to skew focus towards promoting views prompting "top-down" centered control hypotheses. That is, the brain as the end-all and be-all of motor behavior (Chemero 2011). Consequently, this skewing sweeps an important consideration underneath the rug: namely, behavior is not *solely* governed by the CNS. Regularities induced by the body's relationship with its environment allow intelligent control mechanisms to emerge (Pfeifer et al. 2007). Isolating the study of motor control to simple tasks, like reaching or sequence generation, do not easily allow for these perspectives to be taken rigorously enough; put simply, their simplistic nature focused on isolated, adapted movements make studying those properties challenging or impossible.

Creating an approach that sought to rectify these two biases to explain the intelligent feat of

human sensorimotor skill learning is the focus for the work undertaken. Here, this approach critically depends on the following choices heavily dependent between one other:

1. **Experimental Design and Imaging** – Choosing what experimental paradigms and protocols would allow the simultaneous investigation of both behavior and brain. An experimental design needed to be selected such that internally generated activity and task-evoked activity could easily relate to one another. This relationship should reveal properties of learning. Yet, the reason for why this relationship exists needs to be constrained and justified. That is, how is the supposed relationship informed by neuro-physiological mechanisms. Further, this needs to be mindful of resource constraints as far as the number of participants recruited and the number of times a participant could be arranged to be measured. From the image acquisition side, MR imaging protocols that provide stable measurements also need to be considered carefully.

2. **Methodology** – Choosing what kind of data analyses provide the insight needed to support clear conclusions regarding brain organization. Are current methodologies powerful enough or do they need to improve considerably before they are usable in the intended setting? Care must be taken to understand whether current evaluation measures are correct and they are generalizable in other use cases. Further, preference of machine learning over classical statistical approaches often yield higher specificity, robustness, and offer more flexibility over conventional approaches.

3. **Behavior** – Choosing what aspect of human behavior was important to study. Motor skill learning was the chosen focus of this work because of its universality and the analytical rigor it can be characterized by. The chosen skill needed to be difficult enough such that it would require some amount of practice to achieve satisfactory and behavioral effects could be detected. The task should be chosen to be complex enough such that it could resemble highly relevant, everyday behavior. Behavioral research should largely be considered epistemiologically prior to neuroscientific questions. Therefore, it is critical to know what algorithmic properties of this motor task within its environment one would need to consider. Last, is the need for the design of specialized equipment to acquire accurate behavioral assessment in the hostile MR scanner environment.

The rest of this thesis explains what formed the dialog that negotiated these consideration and the results that were consequential of the designed study that followed. An overview of the canonical fMRI experimental paradigms used to study sensorimotor learning is provided in this chapter. Why these two paradigms are important to understand together is discussed

and justification for the methodological focus is presented. Results available in Chapter 2 attempting to link these two separate measurements together with new methodology are detailed. Following Chapter 2's results, Chapter 3 argues why BOLD imaging may provide an opportunity to see plasticity related changes induced by learning novel motor skills. The behavioral background and context of why introducing a new task of virtual paddle juggling into the scanner was a novelty worth pursuing. That is, compared to established motor learning paradigms. The algorithmic properties that underlie the behavior to perform the task are reviewed. Chapter 4 presents results from a study that sought after relating plasticity effects across rest and task fMRI with a single session longitudinal design utilizing virtual paddle juggling. Finally, in Chapter 5, an outlook towards future work where outstanding research questions that extend my thesis results are detailed.

1.2 Task-free versus Task-evoked fMRI - Experimental Paradigms and Methods

How should we understand changes in learning complex motor skills to take place in the brain? In particular, how can fMRI be used to detect the dynamics of learning-induced changes and how should these changes be understood with spontaneous activity versus task-evoked activity? Brain activity measured by fMRI needs to be operationalized such that statistical analysis of brain data can be performed. Typically, this is done via an *encoding* model–a model that specifies how information, either external or internal, is represented by statistical summaries of brain activity (Naselaris et al. 2011). The fMRI measurement paradigms and analysis techniques that do this are reviewed. Together, this understanding formed the basis on how to decide for the appropriate experimental paradigm and, in particular, what motivated the methodological approach.

1.2.1 MR Imaging Paradigms for Examining Motor Skill Learning

MR is a noninvasive mainstay for characterizing human brain changes produced by learning (Draganski & May 2008; Hamaide et al. 2016; Calmels 2020). Use of the imaging technique to investigate learning can largely be divided between functional and structural studies. A fundamental assumption is that functional activity is largely determined by structural properties (Saygin et al. 2012). These two modalities therefore compliment each other and considerable work has focused on understanding their relationship (Honey et al. 2007; Van Den Heuvel et al. 2009).

Structural longitudinal MR studies make up some most recognized early work on human learning (Maguire et al. 2000; Draganski et al. 2004) and are actively used for the design of novel studies (Jillings et al. 2020). They investigate macroscopic gray and white matter density changes that accompany improvement of the task (Taubert et al. 2010; Driemeyer et al. 2008; Scholz et al. 2009; Zatorre et al. 2012). These changes are measured primarily with T1-weighted and diffusion weighted imaging, respectively. The first measures density of gray matter while the later measures directionally dependent changes in water diffusion. Anatomical image contrasts, altered via TR and TE parameters, are acquired via differences in relaxation properties between tissues. Popular techniques used to investigate these two imaging modalities are Voxel-Based-Morphometry (VBM) and changes in Fractional Anistrosophy (FA) or Mean Diffusitivity (MD), respectively. Most of these structural investigations rely on longitudinal designs lasting considerable lengths of time, (Draganski & May 2008), although more recent work claims short-term changes are possible to detect (Keller & Just 2016; Tavor et al. 2020). In any case, functional studies (fMRI) are particularly attractive when short-term learning taking place over a single measurement are studied. Short-term learning measured within a single functional measurement session is the focus of this thesis work.

Functional studies make use of local vascular responses measured via local cerebral blood flow (CBF), cerebral blood volume (CBV), Cerebral metabolic rate of oxygen ($CMRO_2$), or–most commonly–BOLD: blood oxygenation level-dependent, contrast-weighted images. Rapid image acquisition form 3D time-series that allow indirect assessment of the intensity and spatial extent of a neural response. What we measure as signal, physiologically speaking in simple terms, is a **brain activity**. Not long after its introduction, rapid fMRI advances in image acquisition have allowed more detailed functional study of motor skill learning (Kami et al. 1995). Since early studies focusing on task-activity engagement, fMRI continued to develop considerably. Nevertheless, two fMRI experimental paradigms make up the vast majority of investigations into human motor skill learning and are well-established in the community. The first–**task fMRI (tfMRI)**–measures brain activity when engaged under specific manipulations in controlled engagement with the task demands. The second–**resting-state fMRI (rsfMRI)**–investigates brain activity in absence of controlled sensory input and usually with the instruction to "keep awake", "do not think about anything in particular", and/or "visually fixate upon a crosshair display." As mentioned earlier, these two paradigms have largely evolved to be treated separately with little overlap. Yet, both are widely utilized to study learning effects. The subsections below explains why a present dichotomy between these two paradigms exists and why work to bring them closer together

can help resolve some substantial scientific problems in the field.

1.2.1.1 tfMRI

tfMRI is the ubiquitous workhorse paradigm of human cognitive investigations. Conventional tfMRI focuses on the design of experiments allowing for a strict control and isolation of cognitive or behavioral processes under targeted manipulation. The activity of these manipulations are then compared against each other via an encoding model–a model (usually explicitly) specifying the executed behavior or stimulus information presented, e.g., stimulus type, timing, and magnitude on the most rudimentary level. The canonical modeling technique widely used is the General Linear Model (GLM) that produces localized activity maps evoked by the controlled task conditions, their differences between, and overlaps. For review, see (Monti 2011).

Under more detailed investigations, particular stimulus features or behavior may inform a computational model used to infer a hypothesized mechanism of brain function. Coupled with fMRI measurements of brain activity during the task, the output of that computational model allows an investigator to spatially localize a candidate computational function in the brain (O'Doherty et al. 2007). This is done by a regression from BOLD data onto a variable of interest generated by the candidate computational model fit from behavioral data. Additionally, fMRI data can in theory, although extremely challenging in practice[1], provide a means to perform model selection between competing behavioral hypotheses (Wilson & Niv 2015).

Altogether, these kinds of controlled experiments vastly simplify the design and testing of hypotheses for brain activity measured by fMRI. Model fits are averaged across task trials and often among individuals into group representations, i.e., standardized brain maps. This procedure handles complex, noisy signal mixtures of neural information processing by producing concise summaries of volumetric or surface image statistics. For fMRI, where it is necessary to perform major signal averaging due to weak hemodynamics signals corrupted by considerable amounts of noise sources, this eases the burden of designing costly experiments a great deal. Further, modeling procedures and methods are simplified; controlled task-based experiments allow for easy counterbalancing between conditions that make interpreting their subtractions or conjunctions simple. These subtractions and conjunctions are attractive for

---

[1]Statistically significant outcomes from model-based approaches do not necessarily imply that a computational model is specific to that outcome. (See (Peterson & Seger 2018) for overview). That is, it does not eliminate the plausibility of other models under different parameterizations or completely different models (outside the set being considered) altogether. This criticism was initially raised for Dynamic Causal Modeling (DCM) (Lohmann et al. 2012).

investigators because their sensitivity to detect (statistically significant) changes are high. Evidence for this is the approach's massive yield of brain maps reported across the literature (Aue et al. 2009).

Critically, an assumption of this conventional tfMRI approach is that the brain should respond the same to identical repetitions of stimulus presentation or generation of the same behavior. Any signal variance that departs from this assumption is often considered noise. Repeated presentations of the same stimulus are performed for purposes of signal averaging. What this means in practice is that most signal variance observed in fMRI is actually discarded. Yet, this "noise" has been shown to be highly structured, influenced by behavior, and play a role in the formation of task activity (Kisley & Gerstein 1999; Kenet et al. 2003; Bianciardi et al. 2009; Shen 2015).

1.2.1.2   rsfMRI

The tfMRI paradigm described earlier emphasizes that neural information processing takes place in the brain as a unidirectional process. More precisely, the brain is conceptualized as a type of computational input-output or encoding-decoding system integrating distinct brain structures conceptualized as local processing units.[2] An external stimulus, its sensory response, and behavior all follow in one direction. While this framework yields considerable successes, especially among visual system models, it is an oversimplification. We know from studies on the neural circuit level that subcortical input is weak, local connectivity reveals complex excitatory/inhibitory recurrence, and output reflects changes in the balance between the dynamics of excitation and inhibition (Douglas & Martin 2004; Logothetis 2008). Therefore, a simple input-output framework ignores a crucial aspect present in any living animal even before actual birth: internally generated neural activity.

Internally generated neural activity is generally regarded as synonymous with spontaneous activity. It is the activity measured when all other known exogenous influences are controlled or assumed to be negligible. Therefore baseline, intrinsic or default are all compatible terms that have developed alongside this notion.

Presently, we know the brain's support for this type of activity to be essential. Briefly, these reasons, but not limited to, are:

- Roughly 80% of the human brain's metabolic energy is linked to the generation of spontaneous activity associated with neural spiking, cycling of GABA, and glutamate

---

[2]encoding—how stimuli are transformed into patterns of neural activity—and decoding, how neural activity generates behavior.

neurotransmitters (Tomasi et al. 2013). The additional energy needed to generate task-associated activity is, by comparison, extremely small with some estimates placing it at $< 5\%$ (Raichle & Mintun 2006) or $< 10\%$ (Schölvinck et al. 2008).

- Variability of task-evoked cortical responses of a identically and repeatedly presented stimulus can often be as large as the response itself (Arieli et al. 1996).

- Spontaneous activity appears to be highly organized in so-called resting-state networks (RSNs) (Gusnard & Raichle 2001). Since their discovery, these networks have evolved into an elaborate taxotomy of different canonical networks (Uddin et al. 2019).

- The structure of activity is known to vary with salient examples of internal state, e.g., wakefullness, motivation. These examples are known to create strong behavioral effects, altering perception and behavioral performance in simplistic tasks (Bianciardi et al. 2009).

- The structure of these internal states appear to be preserved across different levels of consciouness and different species (Vincent et al. 2007; Margulies et al. 2009).

- Spontaneous activity that dictate these internal states are known to influence sensory processing, cognition, and motor responses (Chaisanguanthum et al. 2014).

- Inter-trial variability appears to be explained from interactions of sensory responses with spontaneous activity (Kisley & Gerstein 1999; Curto et al. 2009).

- Robust internal structure of internal activity is known to be altered by disuse of limbs (Newbold et al. 2020).

- The brain is known to create an internal model of its environment (Berkes et al. 2011) suggesting that spontaneous activity reflects properties of the statistical structure of an animal's natural environment.

Given how ubiquitous, prominent, and costly maintaining this internal, i.e., spontaneous, activity appears to be, its importance cannot be understated. As a consequence, it is not surprising that rsfMRI has emerged as a widespread measurement paradigm for fMRI research. Broadly stated, rsfMRI investigates brain activity in absence of controlled sensory input or behavioral instruction beyond "keep awake." This "rest" state is believed to be the conscious state with lowest overall metabolic demand and its ambiguity is thought to emphasize individual differences (Dubois & Adolphs 2016), although group studies are widely adopted, as well. Coupled with the simplicity that data can be acquired by a relatively stan-

dardized protocol, rsfMRI has become the workhorse of individual subject research. That is, where a shift away from making inferences of general, large populations to predictive models of individuals has occurred. The later holding many implications for the necessary evolution of utilizing fMRI towards any clinical practice (Dubois & Adolphs 2016).

1.3   Methodology - Bridging rsfMRI and tfMRI together

Many task-evoked patterns of brain activity are reflected in activity measured during rsfMRI and are thought to resemble statistical patterns, i.e., functional connectivity, of habitual cortical activations (Strappini et al. 2019) or cognitive processes. How those patterns come about, therefore, needs to be understood in the context of tfMRI or electrophysiological investigations supporting neural origin. Given these two separate fMRI measurement paradigms for making sense of brain activity, how should we understand them together if they are both utilized to understand the same phenomenon of learning? This question is largely a methodological one given the limited ways activity in each measurement can be operationalized to summarize brain activity. In particular, rsfMRI as correlation-based networks and tfMRI as model-based localized activity maps.

1.3.1   Understanding the relationship between rsfMRI and tfMRI

Despite outstanding ambiguity, an accumulation of work so far suggests covarying signal fluctuations measured at rest are useful descriptions of critical, core brain function. Fundamentally, this work largely relies on forming a connection with task-evoked activity. Early work, at least within the cognitive neuroscience community, has had rsfMRI activity's connection to cognition supported by means of reverse inferences to patterns elicited during tfMRI via crude, non-specific investigations. Initially, the original discovery of spontaneous correlations in rsfMRI was prompted by an intial correspondence to task-evoked activation in sensorimotor regions (Biswal et al. 1995). Biswal, et al., after identifying isolated areas elicited during a bilateral finger tapping task found correlated fluctuations of BOLD measured during rest between voxels within that area and the same sensorimotor areas on the opposing brain hemisphere.

This initial discovery later encouraged the use of methodology relying on cortex-wide correlations among groups of brain regions, i.e., networks, especially in low signal frequency regimes (0.01–0.1 Hz). Use of methodology describing networks sparked further investigation that departs from a traditional approach: cataloging isolated cognitive processes into distinct

regions of the brain activity elicited during a given task (Pessoa 2014). Instead, cognitive processes are increasingly thought of as overlapping distributed networks recruited to support many different functions. The last decade has seen a considerable amount of jargon form around these distinct, structured patterns from rsfMRI measurements. Terminology like baseline, spontaneous, intrinsic, and default architecture emphasize that specific patterns found in rsfMRI are broadly fundamental to human cognition processes. The reporting of these patterns, primarily as correlation-based functional networks, i.e., spatially distinct covarying BOLD time-courses, emphasizes phenomenological descriptions of believed information processing roles. For example, since the seminal work of Raiche, the *Default Mode Network (DMN)* entered the neuroscience vernacular and has since defined a whole sub-field in fMRI research (Raichle 2015). The DMN network is believed to emerge from a baseline level of local neuronal activity when the brain is not engaged in a task demanding attention towards the external environment. This association stems from the fact that the DMN appears in regions that become "deactivated" when engaged with the task, a phenomenon that first appeared with reports in visual tfMRI by Shulman and colleagues (Shulman et al. 1997). Since its first discovery, DMN has since been associated with a host of cognitive and neurological characteristics (Uddin et al. 2009).

Because of the alleged relevance of these co-activations, a whole taxonomy has emerged to label alleged canonical brain networks (Uddin et al. 2019). These include fronto-parital control networks and dorsal and ventral attention networks among others (Beckmann et al. 2005). Dense-sampling approaches are yielding even finer details, parcellating these networks into even further sub-networks (Gratton et al. 2018). Again, these descriptions largely stem from the fact rsfMRI is understood to have some semblance to task-evoked activity. That is, correlated fluctuations measured in rsfMRI reveal a mirroring topography of activity elicited during tfMRI paradigms.

Early work has shown these networks appearing in both rest and a limited set of task-based investigations, but more sophisticated statistical methodology and large data sharing initiatives have since emerged (Van Essen et al. 2013). This allows more detailed investigations asking refined questions of whether the structures seen at rest are the same during tasks covering a much broader domain of human cognitive abilities. Seminal work revealed considerable overlap (2/3 of selected ICA network components extracted) between rsfMRI and tfMRI, suggesting that rsfMRI dynamics are utilizing a vital set of functional networks elicited from all possible tasks (Smith et al. 2009). This study was the first to show how an enomorous extent of networks parsed from rsfMRI activity resemble those observed during tfMRI activity.

Further work has reinforced these conclusions. For instance, task-evoked patterns can be modeled with combinations of rsfMRI networks (Bzdok et al. 2016). Cognitive tasks can largely be decomposed into contributions of generalized networks elicited during a rsfMRI paradigm. The extent this is possible is due to rsfMRI networks and task-evoked networks appearing very similar to each other. In fact, functional correlation structure across a variety of cognitive tasks appear to share 80% of their variance with functional correlation structures found in rsfMRI (Cole et al. 2014). These results suggest that the functional networks that appear during tfMRI are shaped mainly by a network architecture present during rsfMRI.

Altogether, work relating the two measurements generally emphasize that task-evoked activity changes from spontaneous activity appear small. Remarkably, recent work relying on dense-sampling approaches suggests that most functional network variability for a group of particpants performing a variety of cognitive tasks across repeated sessions is actually explained best by *individual* factors, *not* cognitive ones (Gratton et al. 2018). Not only does this suggest that task-evoked activity may have less functional relevance than once thought, but more broadly, that cognitive neuroscience has perhaps overly relied on group tfMRI studies to power its conclusions (Satterthwaite et al. 2018).

Lastly, spontaneous activity appears to hold behavioral relevance as well. Electrophysiological recordings from MT+ explain some variance in trial-to-trial variability in behavioral response. (Britten et al. 1996), Additionally, a bi-stable visual stimulus response could be predicted from spontaneous activity alone (Hesselmann et al. 2008). The strength of task-evoked responses has been reported to be predictable from spontaneous activity (Kannurpatti & Biswal 2012). Further fMRI work remarking on the correspondence has relied on rsfMRI predictive ability of individual behavioral measures and cognitive traits (Khosla et al. 2019; Dadi et al. 2019; Pervaiz et al. 2020), observing behavioral state dependencies (Bianciardi et al. 2009) or dependencies to representations of ecological stimuli (Wilf et al. 2017; Kim et al. 2019), or how features within tfMRI persist or change in subsequent resting-state data observations (Hasson et al. 2009; Tambini et al. 2010; Bernardi et al. 2018).

1.4   Mapping rsfMRI to tfMRI using Machine Learning

With large data sharing initiatives like the HCP, machine learning techniques performing individual prediction becomes an increasingly viable and compelling means of uncovering why brain activity between tfMRI and rsfMRI measurement relate to each other. Recent work demonstrates that an individual's collection of functional networks from spontaneous

activity can localize functional activity evoked during that individual's performance of a task (Tavor et al. 2016). Known as "connectome fingerprinting", the approach builds a machine learning model to map rsfMRI features to tfMRI GLM maps. As noted earlier, considerable between-subject variability is present in all fMRI measurement. Group studies problematicaly discard it, although it is a feature that reflects a rich genetic and behavioral basis of variability between individuals. It is not simply a matter of limitations related to spatial normalization of individual subjects into group templates; inter-subject variance appears as a stable characteristic of fMRI data stemming from physiological sources. As a result, considerable differences between individual tfMRI GLM appears. Connectome fingerprinting, by conditioning on data sources apart from tfMRI, seeks to create a prediction that captures that variability reflected in GLM results.

Such predictions are of practical importance. For instance, often functional localizations of cognitive functions cannot be performed or need to be mitigated. This can be due to two reasons. First, in a clinical setting, often patient groups cannot comply with task demands, e.g., attention or movement. This can be due to the fact the patient may be unconscious. However, a localization of brain function may still be extremely useful, especially for pre-surgical planning (Jones et al. 2017). Second, fMRI experiments are costly and time-budgets (independent of financial cost) are often constraining in many studies. Therefore, not needing to perform a localizer run can provide considerable time-savings by enabling a fast, robust estimation of task-evoked activity areas. Lastly, given rsfMRI ubiquitous use and the overall cost and challenge of collecting large datasets required to achieve appropriate statistical power for fMRI (Turner et al. 2018; Lohmann et al. 2018), the limitations of task-based analysis on small datasets may greatly benefit. This may be accomplished by using rsfMRI measurements to generate an informative prior for individual tfMRI brain maps, making inferences of task-based activity more sensitive and specific. Semi-supervised or transfer learning tasks could benefit from learning how to leverage these dependencies further by enabling either higher decoding accuracy or generating encoding models with better specificity (Bartels et al. 2009; Bzdok et al. 2015; Turek et al. 2017; Zhang et al. 2018).

Given its practical benefits, it is not surprising that previous efforts clearly mark the previous decade's literature. Originally, prediction of tfMRI GLM maps first tried to make use of structural MRI features (Saygin et al. 2012). There, tractography of diffusion-weighted imaging (DWI) was used to explore how structural connectivity reflects the architecture of task activations. However, since the relationship between anatomical features and functional ones can be weak, functional features from widepread use of rsfMRI measurement appear particularly attractive. That is, despite spontaneous activity appearing to be heavily

13

constrained by a fixed anatomical basis (Honey et al. 2009).

Evidence for this is available in plenty of recent literature. Langs, et al., show that a multi-atlas label fusion approach yields higher accuracy than a morphological alignment based on anatomical data alone (Langs et al. 2015). Since this largely unnoticed publication, other high profile publications have tackled the problem utilizing rsfMRI data. Tavor, et al., rely on generating rsfMRI features using a Dual Regression ICA approach (Beckmann et al. 2009) coupled with a parcel-averaged linear regression. This approach has since been augmented to include parcel-wise regression methods relying on regularization (Tobyne et al. 2017; Osher et al. 2019). Other approaches have also utilized neural networks (Cohen et al. 2020). Simultaneously, Cole et al. investigating the role rsfMRI networks have to cognitive tfMRI activations, focus on developing an approached called "activity flow" (Cole et al. 2016). Briefly, activity flow is believed to describe the "movement" of activation between spatially disparate brain regions. What parameterizes this "movement" are functional (semi-partial) correlations estimated via rsfMRI data. The approach can be boiled down to a sum of these estimated correlations weighted by task-activation amplitudes. Although not explicitly acknowledged as a kind of connectome fingerprinting, the model sought to provide a more "mechanistic" interpretation over models relying on a more statistical approach to predict tfMRI GLM maps. These methods, among others and newly developed ones relating rsfMRI to tfMRI, make up a great deal of model comparison later found in Chapter 3.

A concentration on this machine learning approach over a classical statistics approach has a few advantages. They are: (1) general-purpose learning algorithms with minimal assumptions regarding the data-generating process can be applied. In turn this allows discovery of complex representations of either rsfMRI or even tfMRI to be arbitrary as long as they preserve spatial information–information that is the basis for the evaluation of these predictions and what largely distinguishes subjects from one another. (2) the task can be formulated as a supervised-learning problem where a clear evaluation of prediction accuracy is possible. Prediction accuracy can easily be summarized on a whole-brain level with a single summary statistic or evaluated on an individual voxel/vertex level. These two types of evaluations might be useful in different contexts. (3) Beyond the utility of predicting task-activation maps for its own sake, more refined hypotheses can be developed and used to describe behavioral factors that may influence prediction accuracy. Machine learning is often criticized as providing solutions that remain difficult to interpret (Gilpin et al. 2018). However, what we understand neurophysiologicaly and behavioraly a priori in the context of sensorimotor learning provides a strong means to resolve this ambiguity when appropriate model comparison is used. Here, we are concerned with the question of why a spatial dependency should

exist between rsfMRI and tfMRI beyond any banal effects related to the measurement or underlying physiology, e.g., anatomical or more MR specific factors. Specifically, we want to disentangle those effects from neural information processing phenomena mediated by behavioral factors associated with learning. Knowing where successful prediction is possible from statistical features associated with effects of learning provide that basis. That is, when appropriate model comparison with naive features that do not capture those learning effects is performed. The next Chapter focuses on how to improve these machine learning models significantly and answers the question of what kind of model evaluation should be used. This work is later motivated to be extended into sensorimotor learning realm into the following chapter.

1.5   Objectives

Work carried out in mapping resting-state features to task-evoked activity sought to accomplish the following objectives:

1. Provide a new, state-of-the-art method for this mapping.

2. Generate a rigorous benchmark comparison of comparable methods that were previously published and examine the influence of focused features modifications on their performance.

3. Investigate the stability of current model evaluation metrics and introduce a spatially resolved metric to map out the predictability of models.

4. Determine whether a behavioral correspondence exists between measures of rsfMRI and tfMRI similarity.

Chapter 2

Jumping over Baselines: Predicting Activation Maps from Resting-state fMRI with New Methods and a Benchmark Examination

2.1   Introduction

The two main experimental fMRI paradigms study the brain in *resting-state (rsfMRI)* and while performing a *task (tfMRI)*. These paradigms were reviewed in the previous chapter. It was observed that brain activity in both share many features that may help to explain brain function (Greicius & Menon 2004; Toro et al. 2008; Smith et al. 2009; Laird et al. 2011; Cole et al. 2014; Gordon et al. 2017; Nickerson 2018; Krienen et al. 2014; Bzdok et al. 2016). Many of these observations show that much of the estimated variance in rsfMRI functional connectivity (FC) appears to be shared with tfMRI activation maps. These observations are often based on group averages. However, averaging across groups destroys relevant information (Stelzer et al. 2014). Therefore, predictions about individual brains are vital for making progress in neuroscience (Gordon et al. 2017). The relationship between rsfMRI and tfMRI for individual subject prediction can be captured by a regression problem, as illustrated in figure 2.1. This topic has been addressed in numerous studies (Langs et al. 2015; Tavor et al. 2016; Cole et al. 2016; Jones et al. 2017; Tobyne et al. 2018; Osher et al. 2019; Cohen et al. 2020; Niu et al. 2020). Here we re-examined methods that address this problem using machine learning techniques with only functional data and attempt to

improve them. That is, learning statistical models mapping rsfMRI and tfMRI data that generalize on unseen test data (individual subjects) (Bzdok & Yeo 2017). Problematically, when considering individual predictions evaluated over the *whole-cortex*, our benchmark comparison shows that previous methods are extremely limited beyond predicting better than a trivial baseline of group averaging. This is alarming. In this chapter we develop a modification of previous methods that allows them to jump over baselines in many cases, though some limitations still exist. These modifications can be briefly summarized as follows: using a regularized regression method that fits and estimates hyperparameters on a *single* vertex or voxel basis. This technique is known from previous fMRI studies (Wehbe et al. 2015), however, has not been used in this context.
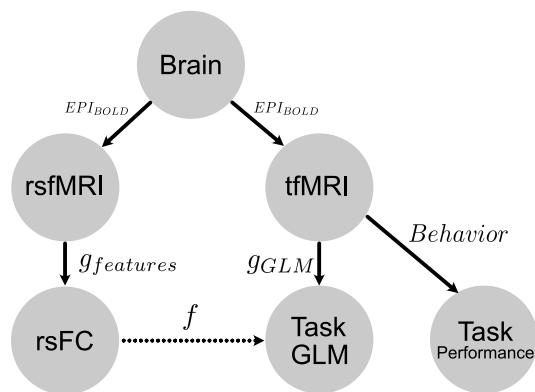
Figure 2.1: A conceptual model of the problem setup and goal. Both rsfMRI and tfMRI measurements are acquired using BOLD echo planar imaging (EPI_{BOLD}) by the Human Connectome Project scanners and acquisition protocols. These reconstructed and processed images from a single subject are mapped by some encoding model: either resting-state functional connectivity features (rsFC) by function $g_{features}$ or tfMRI data mapped to a z-statistical map summarizing task activation by function g_{GLM}. Function f is the model mapping rsfMRI features to task maps. Our goal is to find optimal models $g_{features}$ and f that give the highest performing whole-cortex prediction of task GLM maps (See evaluation section for metrics describing how model performance score is measured). Additionally, activity during the acquisition of tfMRI generates some observed behavior commited during the task. Whether the relative dependence between rsfMRI and tfMRI tell us anything about task behavior is an important question we sought to answer through improving these models.

Therefore, the first aim of the present paper is to demonstrate that the methods we propose are capable of superior prediction. To do so, we provide a benchmark comparison showing how our modifications improve models considerably on a large Human Connectome Project (HCP) dataset. Following these modifications, model predictions achieve above baseline performance for a large number of target contrasts. Notably, these results not only predict individual subject differences, i.e., 'connectome-fingerprints' (Saygin et al. 2012; Finn et al. 2015), as many have previously shown (Tavor et al. 2016; Cole et al. 2016; Jones et al. 2017; Tobyne et al. 2017; Tobyne et al. 2018; Osher et al. 2019; Cohen et al. 2020; Niu et al. 2020); they provide support that whole-cortex prediction by a model exceeds what any kind

of group averaging, i.e., baselines, could achieve–a point we will reiterate the importance of.

Second, to investigate the benefits of the proposed vertex-wise regression, we consider a set of algorithms for feature extraction and prediction, see table 2.1. Besides comparing relevant methods in the literature, we also provide additional insights into which features are actually predictive and discuss other aspects worth investigating. For instance, we give evidence for the relevance of the vertex-wise regularization strategy. Also, we found that widely adopted parcellations surprisingly do not outperform random projections by a considerable margin initially expected for this task.

To arrive at these insights we report additional metrics that we believe should be included in these kinds of studies in the future. That is, in addition to a widely accepted metric evaluating whole-cortex predictions, we report predictive variance explained (R^2 according to sum of squares) on a single vertex level. This examination allowed us to empirically investigate where predictions performed well spatially, explaining why predictions of only a certain number of contrasts perform by a respectable margin above naive baselines.

Finally, recent literature finds correspondence between rest and task activity to be rich in information about individual subject behavior (Schultz & Cole 2016). Following this line, we explore the behavioral relevance of the rest-task dependency found by our best performing method. Namely, we check whether the prediction scores for individual subjects based in rsfMRI carry any information about their behavior during the tfMRI acquisition. We demonstrate how a model's prediction score can be taken as a relative measure of dependency between rest and task measurements. In this way we show that this model may provide information relevant within a behavioural neuroscience context. We also evaluate these behavioral measures relative to a group average baseline. Our results show a compelling behavioral correspondence between resting state and a subject's task performance in certain contrasts. We believe this can drive further progress in the field.

2.2   Materials and methods

We consider fMRI data in the standard format given in CIFTI-greyordinate space. In this study we use data from the Human Connectome Project (HCP) S900 release (Van Essen et al. 2013) and use 100 subjects for training and 100 subjects to make predictions. Here, we consider prediction targets of each subject i to be fixed-effects task GLM maps on the cortical surface $y_i \in \mathbb{R}^{59412}$ defined across 7 different tasks. In total, 47 different GLM contrasts are considered. To model these predictions, we consider methods that first rely on

some feature extraction. For each subject i we consider the data matrix $X_i \in \mathbb{R}^{v \times t}$ where $v = 91282$ is the number of vertices in the CIFTI-greyordinate space and t is the number of samples acquired in time. Further details on the pre-processing of X_i and computation of y_i are found in section 2.2.4.

2.2.1 SOFTWARE IMPLEMENTATION AND USAGE

Python was used for all reported experiments and implementations with the exception of model **GICA-DR-ICA**. This model was implemented in Matlab using code shared from the authors (Tavor et al. 2016). Scikit-learn provided state-of-the-art statistical learning algorithms (http://scikitlearn.org) (Pedregosa et al. 2011). Additional experiments used code modified from the nilearn library for high-dimensional neuroimaging datasets (http://github.com/nilearn/nilearn) (Abraham, Fabian Pedregosa, et al. 2014).
Flatmap cortical visualizations used code modified from (Gao et al. 2015). The neuroinformatics platform that allowed downloading large datasets and a tool for 3D cortical visualizations used software provided by HCP (Marcus et al. 2011). Code will be released following publication.

2.2.2 EVALUATION

Before detailing feature extraction and a new modeling approach, we would like to bring attention to important details regarding how the models are compared against each other. All model evaluation measuring predictive performance is done only on the cortical surface with 59412 vertices within the 100 subject test-set. Individual subject scores were computed as the Pearson correlation score r_i for subject i between prediction $\hat{\mathbf{y}}_i$ and "true" activation map $\mathbf{y}_i$. This image similarity metric is a unit-less measure that provides a concise summary of whether the overall shape of activation prediction is determined to be accurate (Sochat et al. 2015).

This measure alone, however, does not inform us *where* spatially the model is capable of making accurate predictions. For that, we include the predictive R^2 score, providing a measure of how much variance of each vertex is explained by the predictive model. Given test-set predictions at vertex j as $\hat{\mathbf{y}}_j \in \mathbb{R}^{100}$ and "true" activation map $\mathbf{y}_j \in \mathbb{R}^{100}$, this score is computed as

$$R_j^2 := 1 - \frac{||\mathbf{y}_j - \hat{\mathbf{y}}_j||^2}{||\mathbf{y}_j - \bar{\mathbf{y}}_j||^2} \tag{2.1}$$

where $\bar{\mathbf{y}}_j = \sum_i^{100} y_{ij}$ is the subject-wise mean over 100 test-subjects. This score indicates

where and to what extent prediction was possible for each vertex of the fitted model. For whole-cortex comparisons, the weighted average R^2 across the cortical surface was computed. This was done by weighting each vertex R^2 by the variance of the target sample.

Note that we do not report higher intra-subject vs. inter-subject prediction scores as an evaluation criterion as was done in (Tavor et al. 2016). We do not believe this observation is particularly constructive beyond the two evaluation metrics above we use. This position is based on the following observations. We understand intra-subject dependence between separate, spatially normalized whole-brain measurements exists to the extent it allows highly accurate subject identification from both rsfMRI and tfMRI-based measurements (Finn et al. 2015; Byrge & Kennedy 2019). We could expect that an output derived from an arbitrary encoding model of rsfMRI compared to tfMRI activation maps could reveal higher intra-subject correlation than inter-subject, preserving the dependency structure defining rsfMRI and tfMRI are both acquired from the same individual brain. Yet, that prediction can be vastly poorer than a naive, unfitted baseline model in terms of whole-cortex evaluation. Figure 2.2 illustrates that an arbitrary FC encoding of rsfMRI can demonstrate exactly this. A correlation map produced by a random averaging can show higher intra-subject than inter-subject scores to task activation maps clearly marked. This illustrates that inter-subject differences exist despite explaining no variance on a vertex-wise level and vastly underperforming baseline scores. While this observation still reveals individual features unique to the subject are preserved, we hesitate to claim it is evidence of a successful prediction about something unknown.

Instead, we believe it reiterates what we know from the very outset of the problem: both rsfMRI and tfMRI are measured from the same brain. Therefore, we try to place our claims of predictability by emphasizing comparison against models of massive subject averaging. Our goal is that our prediction performance exceed these simple subject averaged baselines across the whole cortex.

Also note that we specifically choose not to evaluate any model performance based on a measure of suprathreshold extent, e.g., thresholded maps and their overlap indices–Jaccard or Dice. We also do not report any qualitative comparisons based on suprathreshold extent as we believe it can be misleading. Dice overlaps and higher intra-subject than inter-subject scores have been widely used in related literature to present support for successful model predictions performance. We believe reliance on these metrics can be highly misleading and have therefore avoided its use. This is illustrated with two examples that highlight this. First, figure 2.2 shows how higher intra-subject than inter-subject scores can be trivially obtained to show individual differences. Second, Dice coefficients have properties that make the metric
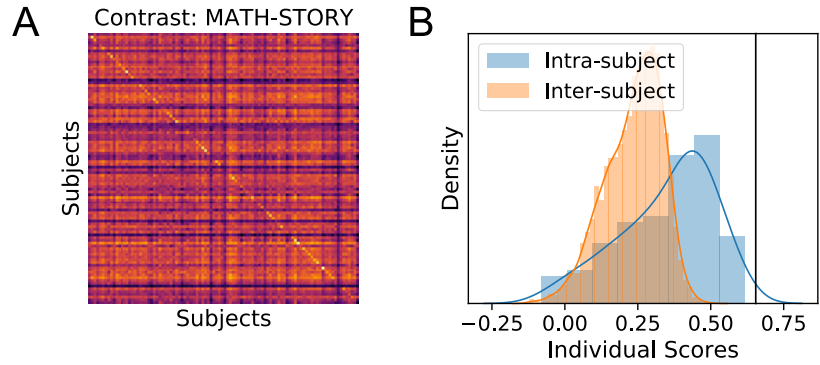
Figure 2.2: Higher intra-subject than inter-subject scores for task activation map prediction can be shown with minimal modeling and requiring no statistical fitting. We illustrate this with an arbitrary example under the Math-Story contrast target. Here, a random projection (see rsfMRI feature extraction section) where A is sampled from $\mathcal{N}(0,1)$ can demonstrate a higher intra-subject than inter-subject scores to a subject's task activation map. (A) shows the subject-wise confusion matrix. (B) shows a density plot of intra-subject vs. inter-subject scores. Brains measured from the same subject with rsfMRI and tfMRI are more similar to each other than different subjects. These scores are still considerably worse than a **Group Mean** baseline marked in the solid black line.

poor for evaluating model performance. This is based on two factors: (1) sensitivity to chosen thresholds that change baseline and fitted models performance differently shown in 2.3 and (2) non-intuitive biases that occur as a result of increasing the number of subjects used calculation of baseline models, lowering model performance, especially at conservative thresholds shown in figure 2.4.
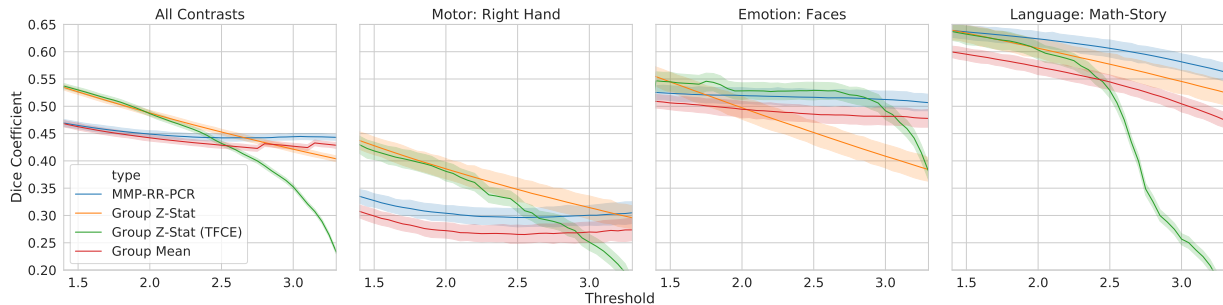
Figure 2.3: Dice score sensitivity to choosen threshold: Single subject GLM maps hold a considerable degree of noise resulting in inflated type-II statistical errors in detecting activation in task-based fMRI. Therefore, many results reported on the vertex/voxel level are often done on a binary active vs. non-active classification basis and are sometimes utilized to evaluate fMRI reproducibility measures. Here, we report on the stability of applying a threshold procedure for the model evaluation. Varying thresholds (Z-scores: 1.4-3.3) were chosen to display the overall sensitivity to results over baseline models: Group Z-stat, Group Z-stat Threshold Free Cluster Extent (TFCE) corrected at Family-wise Error (FWE) $p < 0.05$, and Group Mean. Liberal thresholds show Dice coefficient scores higher for group-based models than a top-performing fitted. Fitted models, according to this evaluation metric, show better performance than **Group** models only when thresholds are increasingly strict.

We found results based on these indices to be highly dependent on their chosen threshold, which acts as a nonlinear transform to spatial maps. Further, we also found that group results are highly dependent on the number of subjects used in a manner that is atypical of increasing sample size influence on model performances.

That is, an increasing number of subjects used for **Group Z-stat** or **Group Z-stat (TFCE)** biases Dice coefficients scores downward when thresholds become conservative, e.g., from Gaussian mixture model thresholding. An empirical demonstration of these influences from chosen thresholds and number of subjects used on predictive versus group averaged models is provided in figures 2.3 and 2.4. These observations together provide the basis why using these two metrics appear inappropriate. Therefore, we do not use them to measure any model performance, which deviates from previous reports. Also, note that such demonstrations undermine confidence of a clinical utility in presurgical planning since in every imaginable case, detecting true positives are of vastly greater importance than false negatives. Therefore, we would advise caution on approaching claims purporting the clinical usefulness of utilizing resting state-based prediction mapping of task activation.
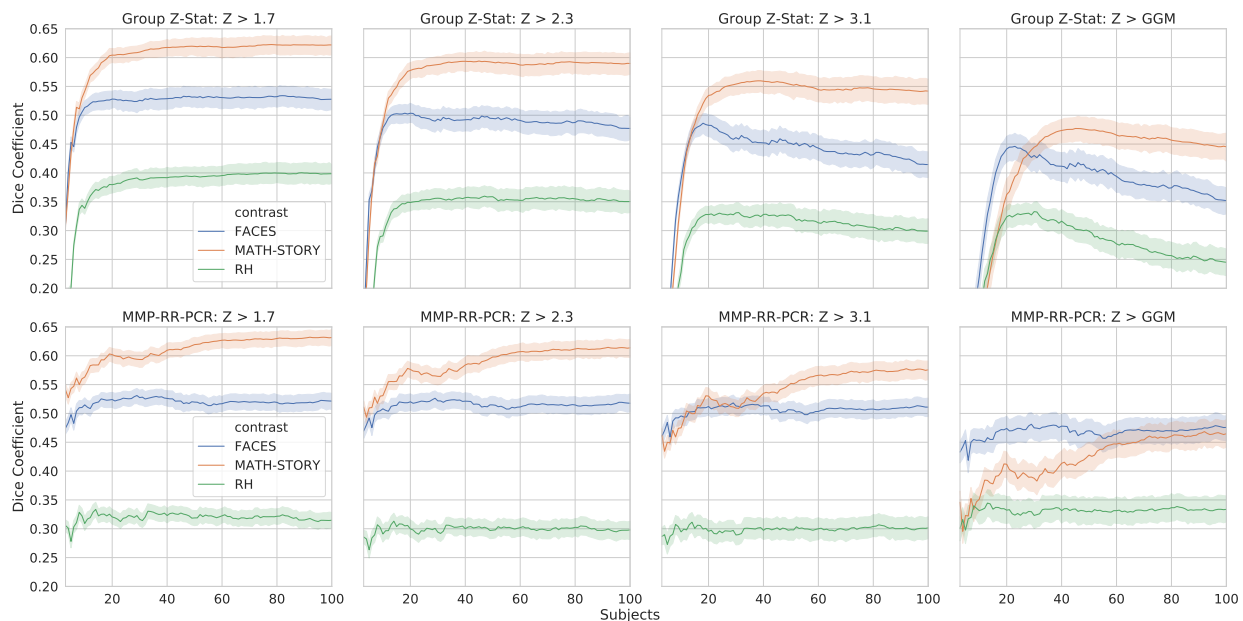
Figure 2.4: Dice score sensitivity to number of training subjects: A comparison between **Group Z-Stat** models (top row) vs. **MMP-RR-PCR** (bottom row) with varying samples (subjects) used to fit the models. Shown are 3 different contrasts used in the main text for subject-wise evaluation. Column plots show increasing (left to right) Z thresholds choosen for the comparison, i.e., 1.7, 2.3, 3.1, and GGM–a threshold calculated from individually fitted Gaussian Gamma Mixture models on actual task activation maps of individual subjects, which tend to be considerably more conservative than $Z > 3.1$. This threshold was the estimated median of the positive Gamma component. Higher thresholds show that an increasing number of subjects used to calculate **Group Z-Stat** models lowers Dice score considerably. For fit model MMP-RR-PCR on contrasts FACES or RH, the fitted contrasts appear to remain flat across all thresholds without considerable and expected increases in Dice score due to increasing sample size.

Model Name	Proposed here	Parcellation - Feature Extraction	Type of Fitting	# of features
MMP-RR-PCR	✓	MMP - Partial Correlations	SV Ridge Regression	379
Rest-Task GICA RR	✓	ICA on Task Data	SV Ridge Regression	80
Rest-Rest GICA RR	✓	ICA on Rest Data	SV Ridge Regression	80
MMP-RR-DR	✓	MMP w/ Dual Regression	SV Ridge Regression	379
MMP-RR	✓	MMP	SV Ridge Regression	379
GPR-RR	✓	Random Projection	SV Ridge Regression	379
AF-Mod	✓	Mean Activation Maps	SV Linear Regression	1
GICA-DR-OLS (Tavor et al. 2016)	✗	ICA w/ Dual Regression	Parcel-wise Linear Regression	50
MMP-ParcelRR (Tobyne et al., 2017)	✗	MMP	Parcel-wise Ridge Regression	360
MMP-OLS	✓	MMP	SV Linear Regression	379
AF (Cole et al., 2016)	✗	Mean Activation Maps	None	∅

Table 2.1: An overview of all methods we compare and benchmark. The names are composed of parts for feature extraction (**MMP**, **PCR**, **GPR**, **GICA**, **AF**) and regression model (RR, OLS), see Methods for details.

2.2.3 MODELING

The subsections below will detail various feature extraction methods used in the benchmark evaluation. For making predictions we are comparing a number of existing methods, as listed in table 2.1.

2.2.3.0.1 Vertex-wise Ridge Regression Model. We propose to use a regression model for *each* vertex j independently, each with its own hyperparameters. We use a ridge regression model fit over all train subjects $n = 100$ as

$$\hat{w}_j := argmin_{w_j}||\mathbf{y}_j - \mathbf{G}_j w_j||_2^2 + \lambda_j||w_j||_2^2 \tag{2.2}$$

where $\mathbf{y}_j$ is an n-dimensional vector of task activation belonging to vertex j (here $j = 1, ..., 59412$) on the cortical surface. $\mathbf{G}_j$ is the $n \times f$ feature matrix of extracted f number of rsfMRI features, as detailed below. Any model making use of this vertex-wise ridge regression is denoted by **RR** in its complete model title. Since the regression model is typically under-determined, regularization is essential for generalization of the model. We chose a quadratic regularization with hyperparameters λ_j controlling the degree of regularization separately for each j vertex. The values λ_j were chosen via a generalized cross-validation procedure over the training-set data (Golub et al. 1979). We suspected that any method offering some degree of shrinkage would be suitable (Wehbe et al. 2015; Tobyne et al. 2018).

2.2.3.1 Baseline Models

Three baseline models were used to judge the actual prediction performance of all models listed in table 2.1. A first and most obvious choice is simply the mean (**Group Mean**) of our targets computed from the training set data. Further, we computed group-level Z-statistics

with multiple comparison correction (**Group Z-Stat (TFCE)**) and without multiple comparison correction (**Group Z-Stat**) for every contrast. **Group Z-Stat (TFCE)** results were only used to investigate the results based on suprathreshold extent. \

Group Z-Stat (TFCE) details. Calculation of **Group Z-Stat (TFCE)** was performed as follows: group-level statistical inference was calculated within CIFTI-greyordinate space form via Permutation Analysis of Linear Models (PALM) (Winkler et al. 2014) that incorporates spatial statistics for Threshold Free Cluster Enhancement (TFCE) (Smith & Nichols 2009). This was done by performing inference separately for each brain hemisphere and CIFTI volumetric component since surface and volume space yield different spatial statistics used by TFCE. Surface meshes to incorporate these spatial statistics were computed as the mean mid-thickness cortical surface across all 200 subjects via the Connectome Workbench toolbox (Project 2019a) **wb_command -surface-vertex-areas** and then merged together **wb_command -metric-merge** and finally averaged **wb_command -metric-reduce**. Group-level one-sample t-tests were computed ignoring intra-subject variance of contrast parameter estimates for all 47 contrasts examined. To facilitate quicker computation times, p-value estimates were computed via a Gamma fit approximation with 500 permutations performed for the null-distribution estimate of each contrast assuming independent and symmetric errors as implemented in PALM (Winkler et al. 2014). Group-level maps were finally transformed to a z-statistic and thresholded at $p < 0.05$ (FWER-corrected for multiple comparisons using TFCE).

Finally, as an additional baseline model, we fit a ridge regression model separately for each surface vertex with 6 anatomical features (**Anatomical RR**). The motivation for including the anatomical baseline stems from speculation that most variance of task-activation shapes can be explained by the subject's anatomical features. These anatomical features are the mean image across the RL-phase and LR-phase (encoded EPI resting-state session-1 runs) and 4 anatomical T1w features extracted from Freesurfer segmentations (recon-all): cortical (quasi) myelin, sulcal depth, curvature, and thickness maps.

2.2.3.2 Resting-state Feature Extraction

All resting-state models we consider rely on some functional covariance-based (FC) feature extraction of resting-state data on a vertex-wise level. For each subject i the normalized data matrix $X_i \in \mathbb{R}^{v \times t}$ is converted into the feature matrix G_i in the general form as:

$$G_i := A^T X_i X_i^T \tag{2.3}$$

where $A \in \mathbb{R}^{v \times f}$. A projects the subject-sample covariance matrix into a lower dimensional space f (number of features). G_i is also known as a "semi-dense connectome." Matrix A is selected either based on predefined Regions of Interest (ROI), e.g. parcellations, on group-based ICA, or on selecting specific features directly, e.g., random projections or mean task activity. Note, that no smoothing of data matrix X_i was applied before any feature computation in any of methods examined.

Multimodal Parcellations. Let us first consider the case of predefined brain regions using Multimodal Parcellations (Glasser et al. 2016) (**MMP**) with $f = 379$ and $A \in \{0, 1\}$. In other words A averages over the activity in spatial regions. As an additional modification we include an additional step of Dual Regression for feature extraction (Nickerson et al. 2017) denoted as **DR**.

ICA-based. For the case of ICA methods we consider the method of computing A via Multi-subject/Group Independent Component Analysis (GICA) for calculating G_i following the algorithm (Canonical ICA) outlined in (Varoquaux et al. 2010). This was done to compare *group maps* extracted either between rest or task. That is, **GICA** features were either derived from rsfMRI or tfMRI data for models **Rest-Rest GICA RR** or **Rest-Task GICA RR**, respectively. In the tfMRI case, separate features were calculated by selecting only 6 of the 7 tfMRI datasets, leaving out the tfMRI measurement of the to-be predicted GLM task contrast. Doing this excludes circularity. These group-level maps were computed over the 100 training subjects. Briefly, the estimation involved a separation of subject-level noise by applying PCA in the time dimension. These subject-level PCs were then concatenated to estimate group-level patterns via Canonical Correlation Analysis (CCA). Group-level PCs were then finally decomposed into group-level independent sources with ICA via FastICA (Hyvarinen 1999). The number of both subject-level and group-level components selected was 80 (Dadi et al. 2019). Note, we did not apply further region extraction from these group-level maps to obtain non-overlapping, parcellations. Hence, the number of features remains at 80.

Activity Flow. A method called "Activity Flow'', **AF** (Cole et al. 2016) uses a group-mean task-activation pattern computed across the training set and uses it directly for prediction for held-out regions of the cortex (as defined by some parcellation) without data-driven fitting. Note, like all other models, we do not perform spatial smoothing on the rsfMRI data. We add a version where this is selected as a single feature used for regression called **AF-Mod**.

Random projections replacing parcels. To assess the impact of the parcellation, we replaced the standard parcellation with a random projection scheme. Random projections are a technique for dimensionality reduction using a random matrix having unit column norms such that the projected lower-dimensional subspace approximates the original distances between data points. Provably, if data points in a vector space are projected onto a randomly selected subspace that is sufficiently large, distances between data points are approximately preserved (Johnson et al. 1986). In our case, A is a randomly generated matrix drawn from $\mathcal{N}(0, 1/f)$ denoted as Gaussian Random Projection (**GRP**).

Principle Component Regression. Functional correlation features extracted from **MMP** models so far do not distinguish between direct and indirect interactions of whole-cortex brain activity to time-dependent signals averaged within parcels. In order to compute features that resemble direct interactions more closely, principle component regression **PCR** is used to compute a semi-partial covariance feature matrix G_i for each subject (Jolliffe 1982; Cole et al. 2016). This was accomplished by masking vertices for exclusion within a crucial area surrounding each parcel. Since neighboring vertices are spatially autocorrelated, this step is essential. In detail, **MMP** partial covariance matrices were computed for each subject by projecting a masked data matrix $\hat{X}_i^k$ for each ROI, $k = \{1, ..., 379\}$ belonging to the **MMP** parcellation and masking all surrounding vertices within a 10mm neighborhood of vertices belonging to the kth ROI. Surface cortical distances were estimated as their geodesic distances on a group-averaged (all 200 subjects) midthickness surface mesh. Subcortical distances were estimated by their Euclidian distance within MNI space. For every masked ROI, 512 principle components (PC) were computed via a randomized singular value decomposition (SVD) (Halko et al. 2011). These selected PC covariates were then regressed using ordinary least squares (OLS) onto the selected k ROI mean signal averaged time-series. Estimated regression coefficients from this regression were then projected back into the original 91282 dimension space of the original data matrix X_i. This together results in a same sized subject feature matrix G_i based on **MMP** as used in other models that only compute covariances.

2.2.3.3 Modified Activity Flow Model

As mentioned above, the Activity Flow model performs no statistical fitting to task activation maps. We include our technique of vertex-wise regression to the **AF** model, denoted as **AF-mod**. In detail it is learning a simple two parameter OLS model fit of $G_i \in \mathbb{R}^{59412}$ to task maps for each surface vertex. This was similar to our other vertex-wise models. Additionally, model **AF-mod** does not perform spatial masking of vertices surrounding the

'to-be' predicted vertex as done in the original Activity Flow conceptualization. We do not perform region or vertex prediction in held-out regions.

2.2.3.4 Remarks on Method Choices

Importantly, we *only* use BOLD data features for all resting-state data model evaluations since this is what underlies our significance claims, deviating from (Tavor et al. 2016). Also, a 100/100 train/test split was used rather than the leave-one-out cross validation employed in (Tavor et al. 2016; Tobyne et al. 2018; Cole et al. 2016). In all cases, all features for each subject were normalized to zero mean and unit norm. Note that we did not seek to use an optimal cross-validation strategy to maximize the performance available on the whole dataset, but provide a robust comparison of generalization performance across models given a large test sample size case. Lastly, due to the enormous computational burden of computing a vertex-wise semi-partial covariance matrix, we do not implement the partial covariance model described in (Cole et al. 2016). To do so would be an enormous computational burden that would require downsampling the data since a PCR would need to be computed at each vertex for each subject. Additionally, downsampling the data would render model comparison unfair between models. All evaluations of all model performances were across the same sized data with no additional smoothing applied. Further details regarding model implementation of the Group ICA dual regression OLS model **GICA-DR-OLS**, vertex-wise Activity Flow **AF**, and a ridge regression model fit over parcellations rather than single vertices/voxels **MMP-ParcelRR** may be found in (Tavor et al. 2016), (Cole et al. 2016), (Tobyne et al. 2018), respectively.

2.2.4 fMRI Data and Processing

All data analyzed in this study is from the Human Connectome Project (HCP) S900 release (Van Essen et al. 2013). To limit a number of covariates that are known to be severe confounds to any of the inter-subject analyses, we selected 200 *unrelated* subjects, i.e., no family relatives, with a T1, T2, complete rsfMRI, complete tfMRI, and physiological data acquired. Age mean = 26.28-29.95, SD = 3.37 Additionally, we selected subjects with functional data reconstructed exclusively with algorithm r227. From these available subjects, a random selection of 100 males and 100 females were made.

The study was performed using data provided by the Human Connectome Project (HCP). All data accessed, downloaded, and used by this study was in accordance with WU-Minn HCP Consortium Open Access Data Use Terms

(https://www.humanconnectome.org/study/hcp-young-adult/document/wu-minn-hcp-consortium-restricted-data-use-terms). The study was performed in agreement with those terms. By agreeing with those use terms, no further ethics approval was required at our local institute to use the data. The HCP project (http://www.humanconnectomeproject.org/) is an open National Institutes of Health (NIH) initiative and received the required ethics approval for data acquisition and public distribution. All subjects who participated gave written, informed consent according to the protocol by the HCP consortium as approved by the Washington University in St. Louis Institutional Review Board (IRB). All human data was acquired in accordance with these experimental procedures adhering to these IRB processes by the HCP. These can be found in further detail (Van Essen et al. 2013).

All results in this chapter are performed on a random train-test split (100/100 subjects) of the 200 selected subjects.

Functional data was acquired with highly accelerated gradient echo type echo-planar imaging (GRE-EPI) in 2 sessions on 2 separate days with 2 different phase encoding directions (left-right and right-left). All four 15 minute resting-state sessions concatenated together were used to derive rsfMRI features. 7 tasks were performed during the task functional acquisition (IDs: emotion, language, motor, social, gambling, relational, working memory). Due to some of the potential benefits offered by particular HCP data acquisition choices, data used for our analyses were exclusively in the standard CIFTI-greyordinate space form. This form allows combined cortical surface and subcortical volume analyses without enormous storage and processing burdens among increases in SNR due to surface smoothing and better cortical fold alignments (Glasser et al. 2013).

Minimal HCP preprocessing steps included motion correction, artifact removal, 200 Hz high-pass filtering, ICA-based denoising, and spatial normalization routines that can be found in full detail in (Glasser et al. 2013). Each measurement had its first 5 repetitions discarded before any local processing. All data prior to being applied in any of the models implemented were demeaned and variance normalized (unit-noise variance) feature-wise.

Prediction targets were fixed-effects (2 Sessions) GLM estimated contrast maps over all 7 tasks with a surface smoothing kernel FWHM of 4mm applied. Fixed-effects GLM results were computed by HCP tfMRI pipelines in CIFTI-greyordinate space and z-transformed (Project 2019b). Following (Tavor et al. 2016), the number of contrasts selected was 47 from these 7 task runs.

A cortical parcellation with 360 regions generated by the work of (Glasser et al. 2016)

was used for the left and right cortical surfaces, and we refer to this parcellation as **MMP** (MultiModal Parcellation). Additionally, for completeness and to utilize the volumetric data component of CIFTI-greyordinate space data for feature extraction, we used an additional 19 sub-cortical regions parcellation given by the HCP release, available at (Project 2019b). This results in a total of 379 regions.

2.2.5 BEHAVIORAL DATA

An assessment of cognitive ability of individual subjects was provided by measures tested during tfMRI acquisition (downloaded at https://db.humanconnectome.org). This is used to understand whether individual predictions scores are related to the amount of correspondence between rest and task. Here, we correlate prediction scores (See Evaluation section) to individual behavior measures of cognitive ability. The cognitive tasks for our analysis are behavioral measurements during working memory, language, and relational processing tasks performed while inside the scanner. Following (Schultz & Cole 2016), these tasks were selected primarily because they fulfill normality assumptions. Additionally, they provide the most complete tasks associated with the contrasts we choose for predictions. Pearson R correlation prediction scores were all Fischer-z transformed across all subjects, a variance-stabilizing transformation, before computing further correlations between the behavioral measures.

2.3 Results
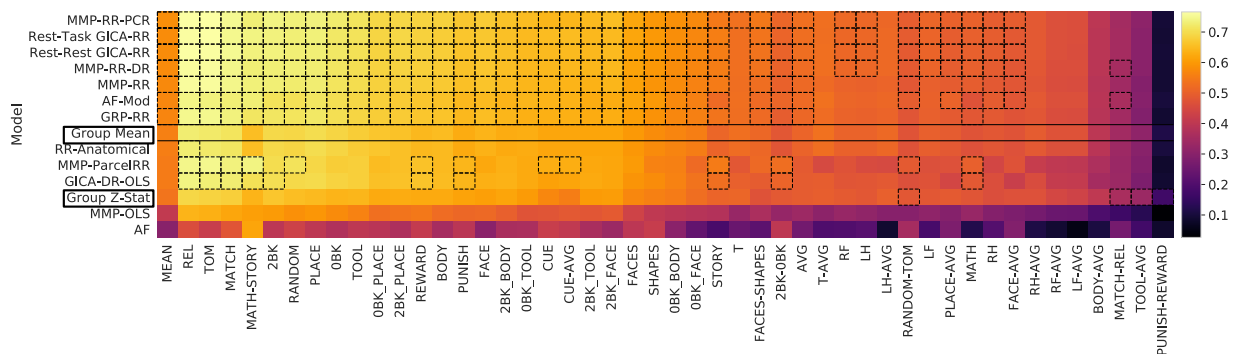
Figure 2.5: Pearson r correlation score benchmark results for 100 subject test set: Colorbar indicates mean r score across all test subjects for given contrast and model. Dashed black boxes indicate where model performance is significantly greater than test-subjects's baseline (mean) model performance (one-sided paired sample t-test, $p < 0.05$, 5000 permutations, Bonferroni corrected across contrast comparisons). Boxes in the left column mark baseline models. Scores are ordered top (best) to bottom (worst) by their subject-wise mean score computed across all 47 different contrasts (left-most column).
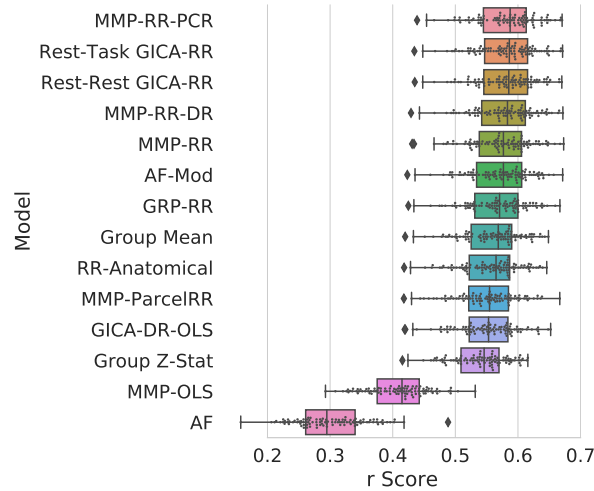
Figure 2.6: Pearson r correlation score results for 100 subject test set averaged across contrasts. Models are ordered top to bottom by score. These scores appear as the first column in figure 2.5. Top performing models perform very similarly between each other. Individual subject scores from the test-set are plotted along with box-whisker plots showing quartiles of prediction score distribution.
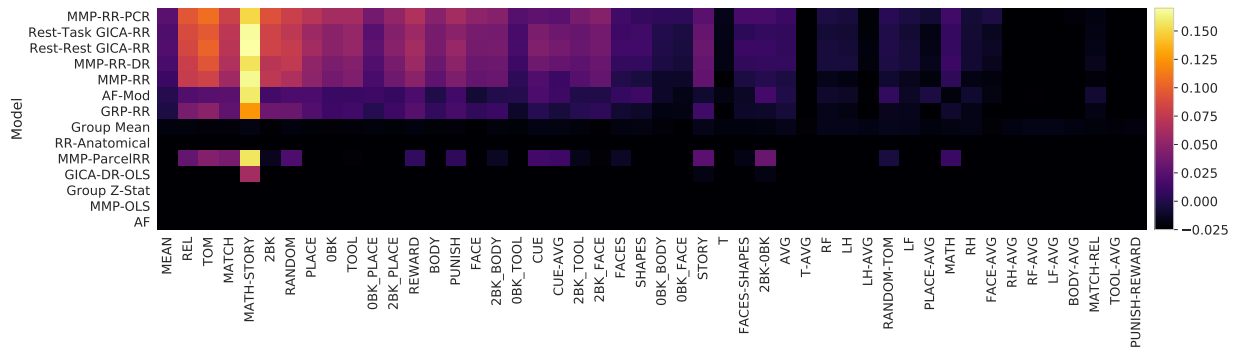
Figure 2.7: Predictive R^2 score benchmark results: scores indicate the mean of cortical surface R^2, see Eq. 2.1, weighted by the variance of each surface vertex. The colorbar indicates this measure. Roughly half of contrast targets have mean cortical R^2 below 0 since predictive R^2 can be arbitrarily negative. Math-Story stands out as the easiest contrast to predict. A discussion providing a reason why is provided in section *Spatially resolved predictability*. Column and row ordering are not sorted by performance and remains identical to figure 2.5. The left most column is the mean score across all contrasts.

2.3.1 Benchmarking: Which Methods Jump over the Baseline?

First, we investigate the accuracy of predictions using the described methods based on Pearson r correlation score for individual subject prediction. We provide a comprehensive performance benchmark comparison with a total of 14 different models. These are compared across the 47 contrast-map targets provided by the HCP S900 dataset. Note that we only focus on model prediction of a single contrast map; this does not leverage any additional information provided by incorporating multiple maps for prediction across subjects. Our

benchmark evaluation compares models using resting-state data against each other and, importantly, against simple baselines models.

This is reported in two figures 2.5, 2.7, summarizing results across the entire cortex according to either Pearson r correlation scores or vertex predictive R^2 scores (Eq. 2.1). The scores displayed in figure 2.5 are provided in the supplements table 2.2. All models were evaluated with the same test-set consisting of 100 subjects. This allows to report statistical significance with a one-sample paired t-test. Importantly, **Group Z-Stat** ($r = 0.540 \pm 0.044$) shows Pearson r correlation score mean performance worse than **Group Mean** ($r = 0.561 \pm 0.047$) for the vast majority of contrasts; only four (three making up the worst performing contrasts) from **Group Z-Stat** performed significantly better than **Group Mean**. Additionally, a model fit only from anatomical features **Anatomical-RR** does not generalize better than **Group Mean** baseline across all but one, the highest scoring contrast (REL).

Therefore, comparisons are made against **Group Mean**, the highest performing baseline model. Many methods, especially from previous approaches, fall short of jumping over this trivial baseline, meaning whole-cortex prediction from the resting state are problematic. Results marked in figure 2.5 by significance boxes reveal that only a limited subset of the 47 contrasts do significantly better than a group mean baseline, **Group Mean**.

2.3.1.1 Improved predictions by vertex-wise models

All methods *with* the proposed vertex-wise fitting procedure demonstrate subject predictions (averaged across contrasts) above the mean baseline prediction (**Group Mean**), figure 2.5 (one-sided paired sample t-test, $p < 0.05$, Bonferroni corrected across all 47 contrasts, 5000 permutations).

Our model **MMP-RR-PCR** yields both the highest mean performance of subject scores averaged over all contrast targets ($r = 0.582 \pm 0.048$) with the highest number of significant prediction performance, see figure 2.5. Additionally, this model holds the highest performance in 31 of the 47 contrasts. However, several other models augmented with our vertex-wise regression method show only slightly worse performance, as figure 2.6 highlights. A direct comparison between the classical way of tuning the ridge regression parameter and our vertex-wised method is seen by comparing **MMP-RR** ($r = 0.574 \pm 0.048$) versus **MMP-ParcelRR** ($r = 0.550 \pm 0.049$), showing a significant gain (one-sided paired sample t-test, $p < 0.001, t = 15.77$, 5000 permutations).

To understand the importance of regularization, we can compare **MMP-RR** ($r = 0.574 \pm 0.048$) and **MMP-OLS** ($r = 0.409 \pm 0.047$), where the latter only relies on ordinary least

squares fits. This drastic performance difference shows that regularization is essential for successful generalization when the number of rsfMRI features is very large. However, a complex model is not necessarily needed for successful prediction; Model **AF-mod** ($r = 0.571 \pm 0.049$) generalizes comparatively very well and has proven to be one of the best performing models despite its simplicity. From our analysis, we expect many methods with some degree of shrinkage would reveal comparable performance when trained on a single-vertex level (Wehbe et al. 2015; Tobyne et al. 2018).

2.3.1.2 Effects of Feature Extraction and Parcellation

We investigate the effect of various feature extraction strategies for determining A in eq. 4.3. First, A derived from task **Rest-Task GICA** data yields only a very small improvement over model **Rest-Rest GICA** derived only from resting-state data, see table 2.2. This motivated us to investigate other effects of selecting A. Specifically, we replaced the expert-based parcellation **MMP** with a random projection A. Again, the advantage of a expert-based parcellation over a random projection is surprisingly small: **GPR-RR** $r = 0.568 \pm 0.048$ vs. **MMP-RR** $r = 0.574 \pm 0.048$. This result suggests that in many cases random projections for generating features appears to be sufficient. It simply provides a means of performing dimensionality reduction akin to perhaps any arbitrary parcellation scheme, an observation consistent with (Tobyne et al. 2018). One crucial point, however, is that the random projections we employ does not take advantage of local-signal-averaging offered by continuous parcellations that are continuity regions. Lastly, we investigate whether deriving more subject specific features via dual regression yielded any appreciable improvement. Model **MMP-RR-DR** over **MMP-RR** shows a statistically significant, yet small, improvement over subject predictions averaged across contrasts (one-sided paired sample t-test, $p < 0.001$). For small sample sizes, however, the use of dual regression appears to be promising, see supplementary figure 2.14.

2.3.2 Predictive R^2 Evaluation

In addition to Pearson r correlation scores, we examine the variance explained on a vertex level (equation 2.1) evaluated on the same test set. This evaluation is summarized in figure 2.7 and provides a complementary measure of prediction performance. The scores displayed in figure 2.7 are provided in the supplements table 2.3. To quantify one number per contrast we report the variance-weighted average of the R^2 scores across the cortical surface.
This number is color-coded in figure 2.7 and quantifies to which degree and in which contrasts predictions about individuals can be made. Models and contrast targets with a positive

R^2 aligns well with the ordering of previous figure 2.5 results and supports how the use of single-vertex regression based methods yields a considerable performance boost and valuable predictions. Nevertheless, figure 2.7 emphasizes that it is only roughly half of the contrast targets that show considerable predictability.

2.3.2.1 Spatially resolved predictability
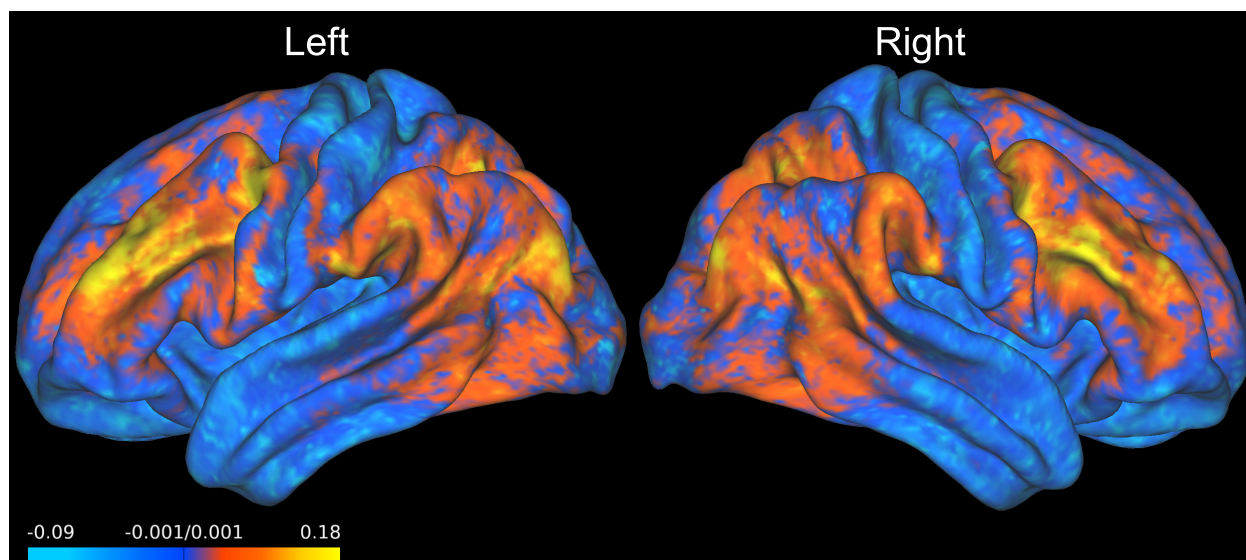
Figure 2.8: Mean R^2 Score of MMP-RR-PCR across all contrast targets. Plotted are the R^2 values averaged over the 100 test subjects. This is rendered on a 200 subject averaged midthickness surface map of left and right cortical hemispheres. Positive values (red and yellow) indicate where prediction is possible. Note that prediction accuracy is best outside the primary sensory regions.

Figure 2.7 shows considerable variability between predictive performance of certain contrasts. This can be explained due to the fact that only certain regions of the cortex drive a model's prediction ability above the baseline. This becomes clearer with an investigation of where on the cortical surface we observe positive R^2 values. To report this concisely, we render the cortical surface with a mean averaged R^2 score across the 47 contrasts of model **MMP-RR-PCR** in figure 2.8. An additional plot showing mean averaged R^2 across each task category separately is shown in figure 2.9.

The surface plot reveals that only a limited subset of vertices lying outside of the primary-sensory regions can explain the 100 test-sample variance. These remain confined within the association cortex where most inter-subject variability of rsfMRI functional connectivity lies (Mueller et al. 2013). Regions of high inter-subject variability as measured by either rsfMRI features, task activation maps, or sulcal depth of a subject's brain anatomy are associated with the predictability, see figure 2.17. This outlines that regions where subject differences
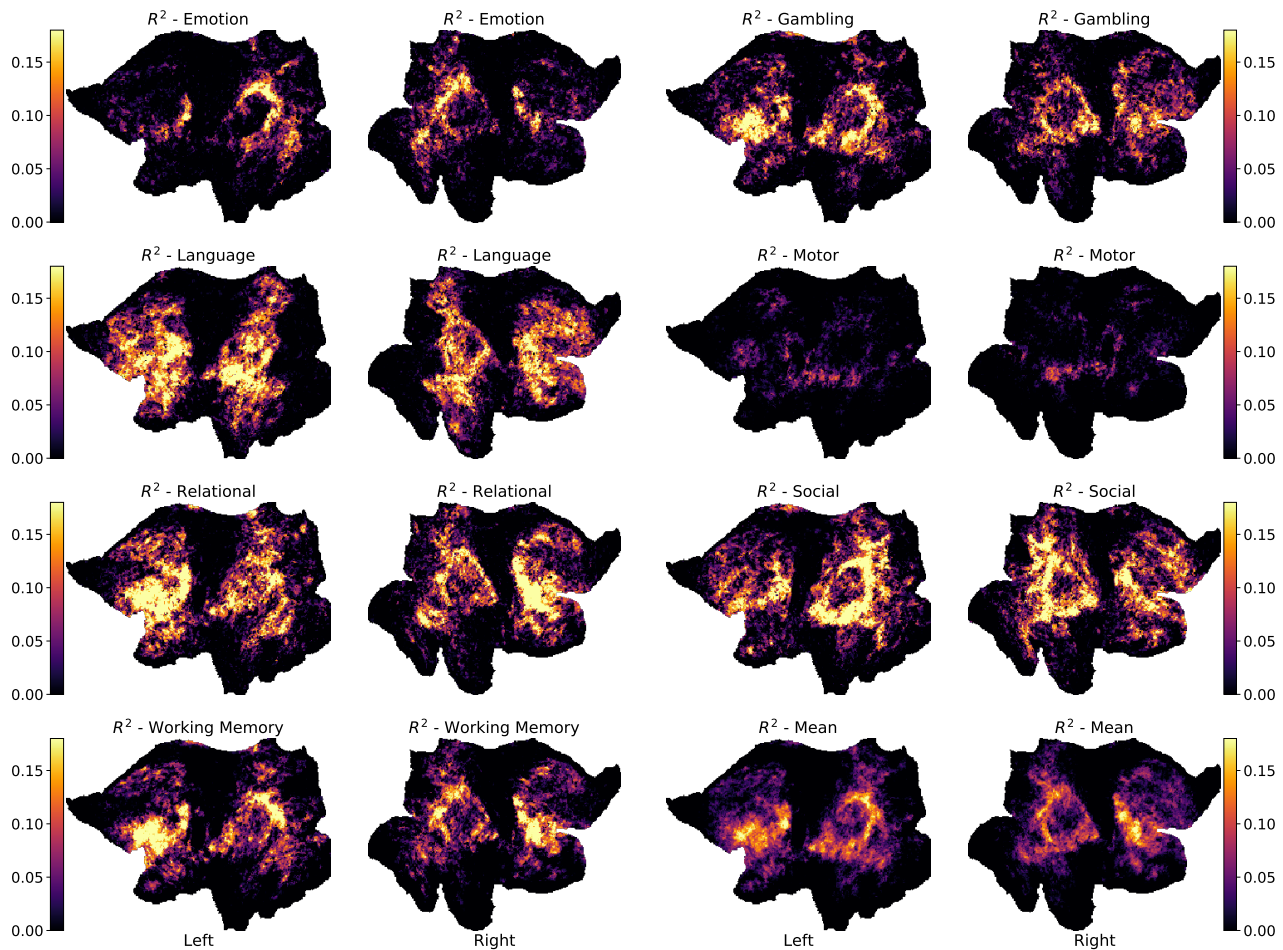
Figure 2.9: Separate test-sample R^2 of model **MMP-RR-PCR** evaluations over task categories: All 47 contrast targets belong to 7 different task categories: Emotion, Gambling, Language, Motor, Relational, Social, Working Memory. Each R^2 evaluation is averaged over contrasts only belonging to its category and plotted in the figure. The test-sample mean is plotted also for convenience. A general pattern is clear across all tasks.

in the cortical functional anatomy are highest are the regions where subject rsfMRI features or task activations also differentiate themselves the most. Supplementary figure 2.18 shows this spatially in flatmap visualization.

To give a better empirical characterization of the spatial dependency of model parameters and prediction quality, we report several metrics per vertex for the **MMP-RR-PCR** model. For visualization we use flatmap cortical projections of the entire cortex, as shown in figure

2.10. We consider the root mean square errors (RMSE) in figure 2.10(A) and see that the highest RMSE appears primarily concentrated around the visual cortex. The vertex-wise strength of regularization λ determined via cross-validation over the training-set is shown in figure 2.10(B). Strong regularization is employed in primary-sensory regions where predictions perform very poorly.

The optimal regularization is inversely proportional to the explained variance shown in figure 2.10(C,D). We show R^2 on the training subjects (C) and on the 100 test-subjects (D).
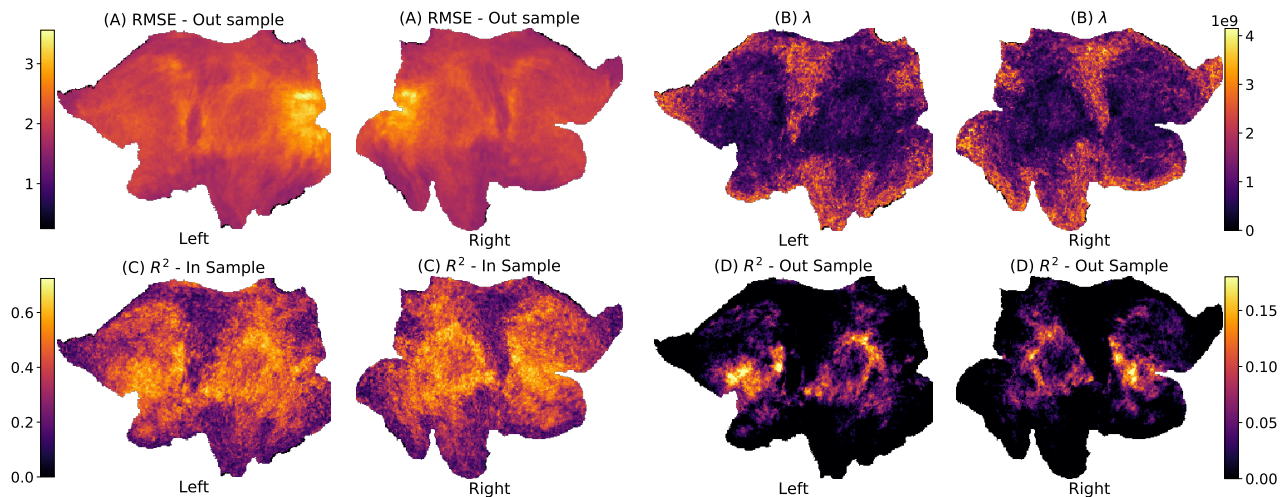
Figure 2.10: Flatmap cortical surface projections of MMP-RR-PCR model fits. (A) Root Mean Square Error (RMSE), (B) Degree of regularization λ in model fit, (C) $R^2 - 100$ train subjects (D) $R^2 - 100$ test subjects. Train and test R^2 show consistent patterns between each other. λ shows how regularization is inversely related to the method's ability to predict (R^2).

2.4 How many subjects are needed?

To examine top performing models closer and according to their capacity, we investigate the impact the number of training samples on 4 of the best models (**MMP-RR-PCR**, **Rest-Task GICA-RR**, **MMP-RR-DR**, **AF-Mod**) as defined by their median contrast score (left most column in figure 2.5. We included two baseline models **Mean (Baseline)**, **Group Z-stat** for comparison. These models were all evaluated on the 100 subject test set. 3 contrast targets were arbitrarily chosen because of their poor, mediocre, good performance as contrasts Motor–Right Hand, Emotion–Faces, Language–Math-Story, respectively. Pearson r correlation scores and predictive R^2 score with respect to the number of subjects (3-100) are reported in figure 2.11 and figure 2.12, respectively.

Figure 2.11: Subject-wise Pearson r score benchmark results for 3 selected (poor, mediocre, good), 4 high performing models (**MMP-RR-PCR**, **Rest-Task GICA-RR**, **MMP-RR-DR**, **AF-Mod**) and two baseline models (**Mean (Baseline)**, **Group Z-stat**). Poor (left): Motor-Right Hand; Mediocre (middle): Emotion: Faces; Good (right): Language: Math-Story. Experiment included 3-100 subjects for training. The group z-statistic baseline results are considerably worse than the group mean baseline. The 4 models largely resemble each other's performance when sample sizes increase past 40 subjects.

All curves of model performance with respect to the number of samples follow typical generalization curves, i.e., an inverse power law, where a rapid increase is seen to a slow saturation when sample size increases (Cortes et al. 1994). As shown in both figure 2.11 and figure 2.5, **Group Z-stats** consistently underperforms its **Group Mean** counterpart by a considerable margin, especially at lower sample sizes. Top performing models largely yield the same performance as the training set increases above 40 subjects.

2.4.1 Additional Subject-wise Evaluation

To discriminate more closely between models in the subject-wise evaluation, the plot figure 2.12 offers a clearer view offered by the R^2 evaluation metric. However, in order to better differentiate between them relative to their sample sizes, the fraction of cortical surfaces having an R^2 above some given threshold is plotted. 3 different thresholds are included to provide the greatest discriminability between models at thresholds $0, 0.1, 0.2$. The sensitivity of the vertex-wise R^2 score from the chosen threshold is plotted in supplementary figure 2.13. For the 3 contrasts examined, at lowest thresholds, model **AF-mod** shows greatest fractional R^2 above 0 for all 3 contrasts while **MMP-RR-PCR** shows greatest fraction R^2 for at least 2/3 contrasts, especially at the highest threshold 0.2. Differences between these models are seen depending on number of training subjects, threshold selected, and contrast examined.

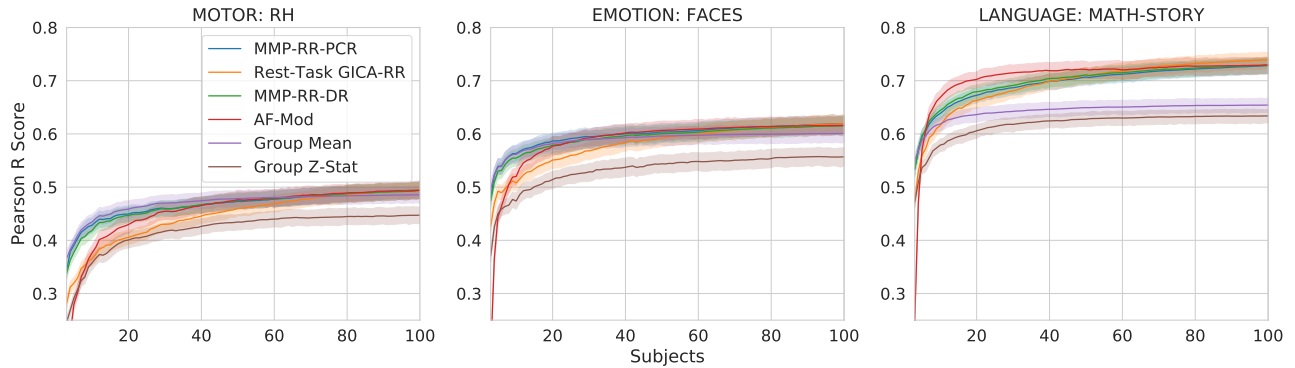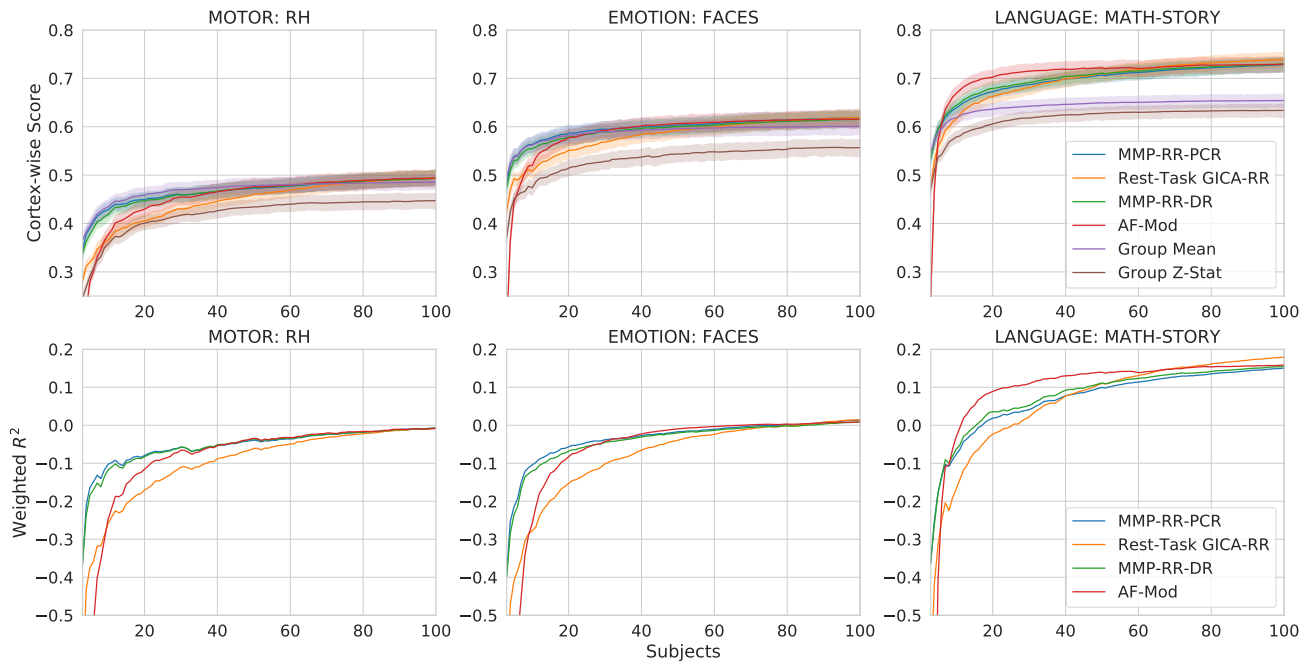Figure 2.12: Top row: Subject-wise Pearson r score benchmark results for 3 selected (poor, mediocre, good), 4 high performing models (**MMP-RR-PCR**, **Rest-Task GICA-RR**, **MMP-RR-DR**, **AF-Mod**) and two baseline models (**Mean (Baseline)**, **Group Z-stat**). Poor (left): Motor-Right Hand; Mediocre (middle): Emotion: Faces; Good (right): Language: Math-Story. Experiment included 3-100 subjects for training. The group z-statistic baseline results are considerably worse than the group mean baseline. The 4 models largely resemble each other's performance when sample sizes increase past 40 subjects.

Bottom row: The weighted mean cortical surface R^2 for four top performing models as a function of number of samples (3-100) used for training. As training samples approach over 80 samples, R^2 largely becomes indiscernible between the 4 models in these contrasts.

2.5 Behavioral Results

Prediction scores may provide a powerful means of summarizing rest-task dependency. We therefore hypothesized that prediction scores may be a means for discriminating behaviorally relevant information about the task performed. It was previously speculated that the degree to which brain activity departs from rest may provide information about individual behavioral performance (Schultz & Cole 2016). Within the network neuroscience community, this phenomenon is recognized as reconfiguration efficiency: high-performing individuals may have brain connectivity that more efficiently updates to the task at hand by not having to produce greater changes in a task functional network organization required to perform the task.

We therefore speculated that if our resting-task model performance for individual subjects
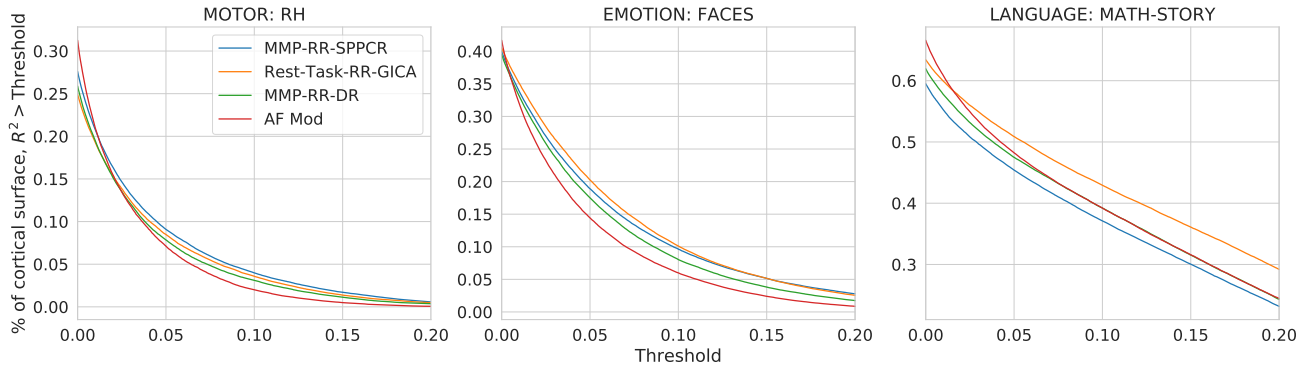
Figure 2.13: Vertex-wise R^2 score sensitivity at 100 training samples: scores indicate the fraction of cortical surface having an R^2 score above a given threshold plotted on the x-axis for given contrast and model.
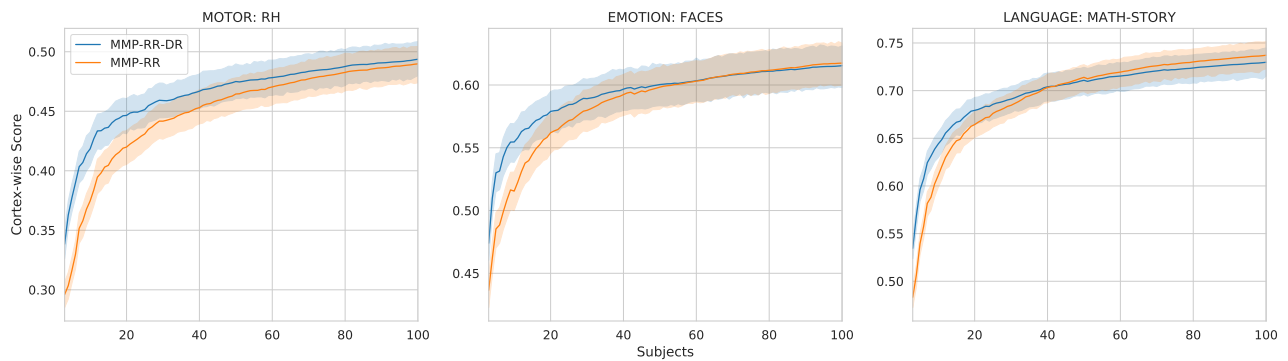
Figure 2.14: **MMP-RR-DR** and **MMP-RR** subject-wise comparison. Applying Dual Regression for feature extraction yielded better vertex-wise and cortex-wise scores for a smaller number of training samples, around less than 40. However, the difference between the two models were negligible when training samples increased to 100. We speculate that dual regression might offer the most benefits with small subject sizes.

could be taken as a relative measure of rest-task dependence, we would see a clear pattern of higher behavioral performance correlating with higher tfMRI prediction scores.

To test this idea, we turn to three behavioral measures of general cognitive ability from Human Connectome Data measured during tfMRI acquistion: working memory, language, and reasoning task. We selected contrasts 2BK-0BK, Math-Story, and Match-Relation since they provided the most general and complete summary of the task and its behavioral data. To see whether prediction scores corresponded to task performance of individuals, on the 100-subject test set we calculate the correlation of individual subject Fischer-Z transformed Pearson r correlation prediction scores to subject task accuracy. This marks whether individual differences in prediction scores correspond to individual differences in behavioral task accuracy. 20 random train/test permutations of 100 train, 100 test subject sizes on the original 200 subject dataset were fit across the models investigated in the subject-wise investigation. Additionally, similar to figure 2.11, we also fit the model **MMP-RR-PCR** from these results for the three selected contrasts under 20 permutations train/test splits with increasing sample sizes (3-100 subjects) and expected that these averaged performance evaluation curves would follow typical generalization curves.

To accommodate that the 20 training and testing permutations were not independent from each other, statistical comparisons between models were made using a corrected resampled t-test (Nadeau & Bengio 2000).

Our results demonstrate that Pearson r correlation prediction scores provide an indicative relative measure of rest-task correspondence to the behavioral task accuracies measured during the performance of these tasks, figure 2.15. All predictive models provide statistically significant results over the baseline for contrast Math-Story (one-sided corrected resampled t-test, Fisher-z tranformed r, dof=19, $p < 0.01$). Mean correlations over 20 train/test permutations for model **MMP-RR-PCR** compared to **Group Mean** was $r = 0.26 \pm 0.05$ versus $r = 0.20 \pm 0.05$, respectively. Models **MMP-RR-PCR**, **Rest-Task GICA-RR**, **MMP-RR** provided statistically significant results over baseline for contrast 2BK-0BK (one-sided corrected resampled t-test, Fisher-z tranformed r, dof=19, $p < 0.05$). The mean correlation for contrast 2BK-0BK over 20 train/test permutations for model **MMP-RR-PCR** was $r = 0.67 \pm 0.07$ versus **Group Mean** at $r = 0.66 \pm 0.07$. However, importantly, in one out of the three contrasts (Match-Relation), no predictive model provides any added benefit over a simple correlation to mean activation (**Group Mean**). That is, despite having strong correlations of $r = 0.40 \pm 0.1$. A plot of individual scores for one permutation (original subject test set) is shown in supplementary figure 2.16 as an illustration of these strong, statistically significant correlations.
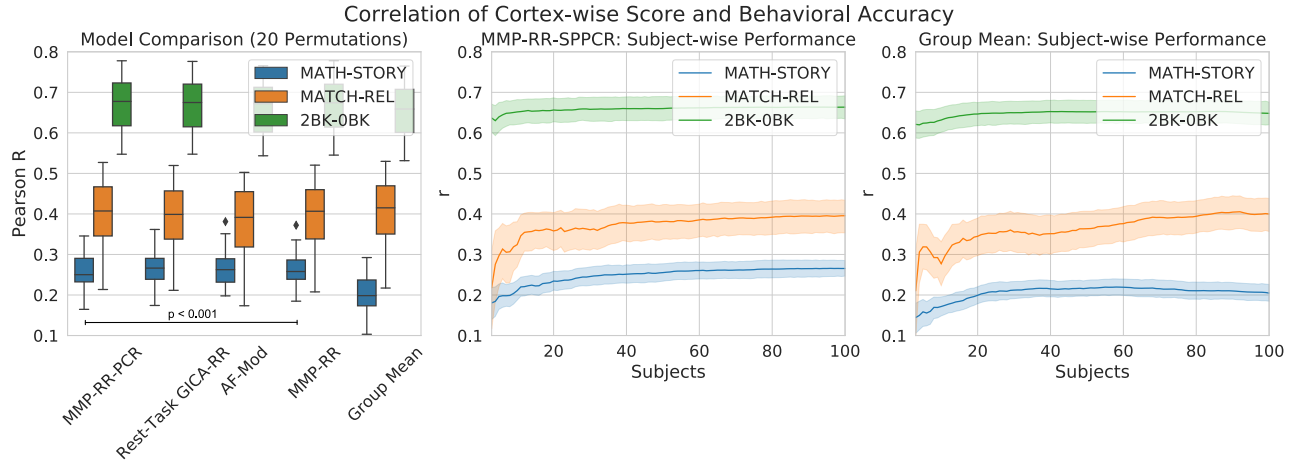
Figure 2.15: Correlation of Cortex-wise Score and Behavioral Task Accuracy. (A) Comparison of different models investigated in detail shown in figure 2.12. Only Math-Story and 2BK-0BK showed significant correlations using a one-sided corrected resampled t-test compared to **Group Mean** (significance marked in by *). (B) Subject-wise comparison of correlation between model **MMP-RR-PCR** prediction score and behavioral task performance for 20 permutation depending on the training set size (3-100). (C) For comparison, same as B, but for the **Group Mean** baseline model.

2.6   Discussion

Motivated by recent progress in establishing a stronger link between spontaneous and task-evoked activity, we examine the problem of mapping rsfMRI measurements to patterns of activity elicited during tfMRI-based experimental paradigms in individual subjects. We show additional evidence that it is indeed possible to predict task activity maps from patterns of rsfMRI FC, as previously reported (Langs et al. 2015; Tavor et al. 2016; Cole et al. 2016; Jones et al. 2017; Tobyne et al. 2017; Tobyne et al. 2018; Osher et al. 2019; Cohen et al. 2020; Niu et al. 2020). However, we emphasized early on that observing higher intra-subject prediction scores compared to inter-subject scores was not useful observation we believed provided informative predictions–they needed perform better than what any naive group averaging could predict on the cortical surface. Our investigation showed that group averaging provided a surprisingly strong baseline for whole-cortex predictions. Results justify selecting group averaging offered by **Group Mean** as a suitable baseline model. This was because it provides substantially higher scores than its alternative **Group Z-Stat**; group Z-statistics were shown to consistently, regardless of sample size, perform below **Group Mean** under nearly all contrast targets. We therefore evaluated all results against the highest performing baseline–**Group Mean**.

Given this appropriate group-averaged baseline model, an examination of previous methods in our benchmark show they did not demonstrate satisfactory whole-cortex prediction scores
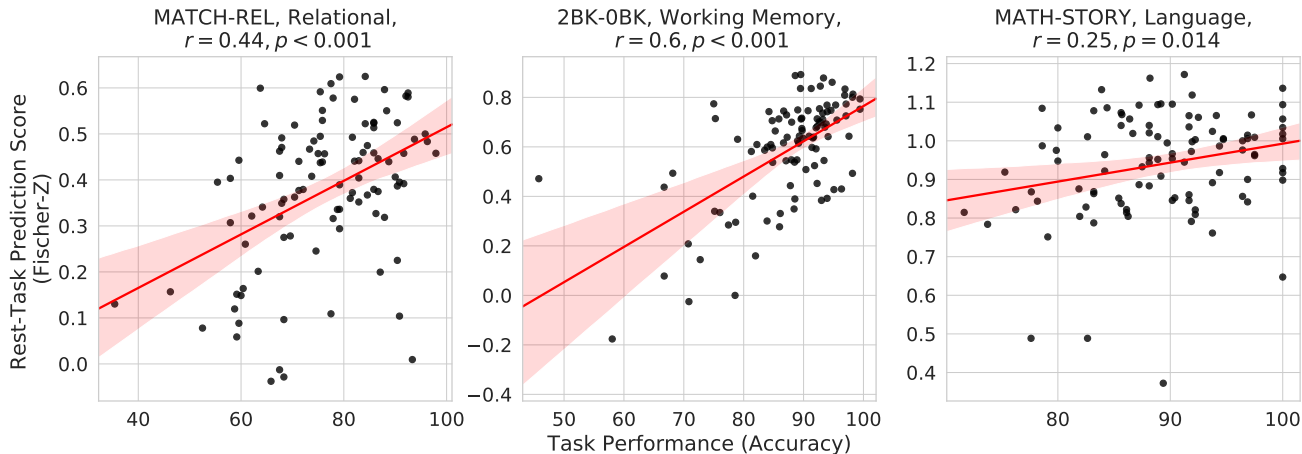
Figure 2.16: Cortex-wise predictions scores of model **MMP-RR-PCR** plotted against behavior task accurary committed during the acquisition of the task. Three contrasts were chosen; Working Memory: 2BK-0BK, Relational: Match-Rel, and Language: Math-Story. Strong correlations are seen for all 3 contrasts.

with a considerable number of contrast targets being outperformed by the baseline. To remedy this problem, i.e., to jump over the baseline, we introduced a simple modification to the fitting procedure: a vertex-wise selection of hyper-parameters. According to our benchmarks, models fit in this manner provide the most powerful means to tackle the problem of predicting tfMRI GLM maps from rsfMRI data we are aware of.

Nevertheless, they also still highlight that in many cases, given the diversity of contrast targets examined, the best performing model we introduced are still modest in their prediction ability with even 100 training samples (subjects).

The considerable variability in prediction scores visible across the 47 contrast maps for all models motivated us to give a better empirical characterization of how this is reflected in model performance. An inspection of the cortical surface areas that have an explainable variance on a vertex-wise level reveals a consistent pattern: primary-sensory regions show little explainable inter-subject variance (R^2) whereas association cortical regions show considerably better predictability. So far, no method appears to be able to explain inter-subject variance within primary sensory regions, as evidenced by strong negative predictive R^2 scores in those locations, figures 2.8 and 2.10. As a further empirical confirmation of these patterns, estimates of inter-subject variability in either FC features or GLM maps are R^2 provided in figure 2.17. Additionally, we also observe these patterns by investigating how the strength of regularization was inversely related to how well the model performs. Together, these observations reinforce earlier work noting association cortex areas hold distributed networks while primary-sensorimotor areas are much more stereotypical across subjects resulting in worse predictions.

Ultimately, our work aims to find which predictions are informative so we could use it to formulate hypotheses asking what behavior or cognitive factors may influence it. That is, the correspondence between rest and task states and how that might reveal information about individual subjects. Seeking to ground this work into a behaviorally relevant context, we considered the question of whether prediction scores of individual subjects provided a means of summarizing rest-task dependence that could inform behaviorally relevant neuroscientific questions given our best performing models. Indeed, the strength of correlation between prediction scores of a given contrast and its corresponding behavioral task accuracy suggests that this prediction score may be taken as a relative measure of dependence between rest and task activity. However, this is not without caveat that places us back to comparing against Group averaged models from the beginning; it is only the case when predictions are considerably above baseline performance we see the utility of performing these model fits. Considerable correlation between the naive model's prediction of **Group Mean** and individual behavioral performance was present for 2 of the 3 contrasts we examined in this way. This fact reinforces our motivation from the very outset of this problem: to create and utilize a method to perform above naive, baseline models. Results shown in our behavior evaluations reiterate this importance.

Our vertex-wise evaluation based on predictive R^2 reveals that considerable performance improvement is still needed to explain variance within primary sensorimotor regions. On speculating how to further improve the methods, we suspect that further significant gains in performance may be obtained from projecting individual FC data into common/shared response spaces via shared response modeling or hyperalignment (Conroy et al. 2009; Conroy et al. 2013; Langs et al. 2015; Guntupalli et al. 2018; Richard et al. 2020). This could provide a means for capturing a substantial amount of inter-subject variance in a manner similar to a dual regression approach (Nickerson et al. 2017). Additionally, separate evaluations reveal that the closer the extracted features are to task-related activity, the better cortex-wise prediction scores are, figure 2.19. We would therefore expect that the use of naturalistic stimuli over rsfMRI could substantially aid over the use of rsfMRI data and would additionally provide the means for additional shared response modeling approach assumptions (Finn et al. 2020).

Some limitations have come to our attention that are important to clearly state when interpreting our results. First and foremost, should rsfMRI fluctuation amplitudes depend on other factors completely unrelated to cortical computations that generate the spatial dependencies we observe with connectome fingerprinting, this would show up in these prediction results. It would additionally confuse interpretation of behavior factors (Bijsterbosch et al.
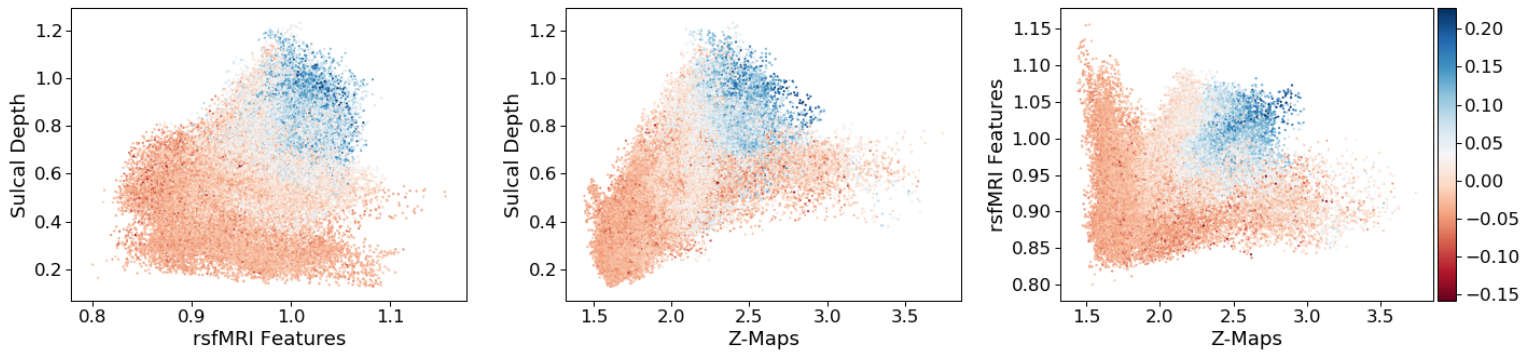
Figure 2.17: Inter-subject variability of activation maps (Z-Maps), sulcal depth, or **MMP-RR-PCR** rsfMRI features and how they relate to measured R^2 and each other plotted as a 3D scatter plot where each point represents a cortical vertex. The colormap represents a vertices' mean R^2 computed across all 47 contrasts. Inter-subject variability of features, sulcal depth, or Z-maps are computed as the vertex-wise standard deviations. rsfMRI features and Z-Maps are averaged across all features (379) or contrasts (47). This plot shows sulcul depth and rsfMRI functional correlational features are strongly correlated to one another. Unsurprisingly, the model's prediction ability as measured by R^2 on a vertex-wise basis are concentrated around the point-cloud mass where inter-subject variability between the two factors are highest (upper right-most areas of the plots).
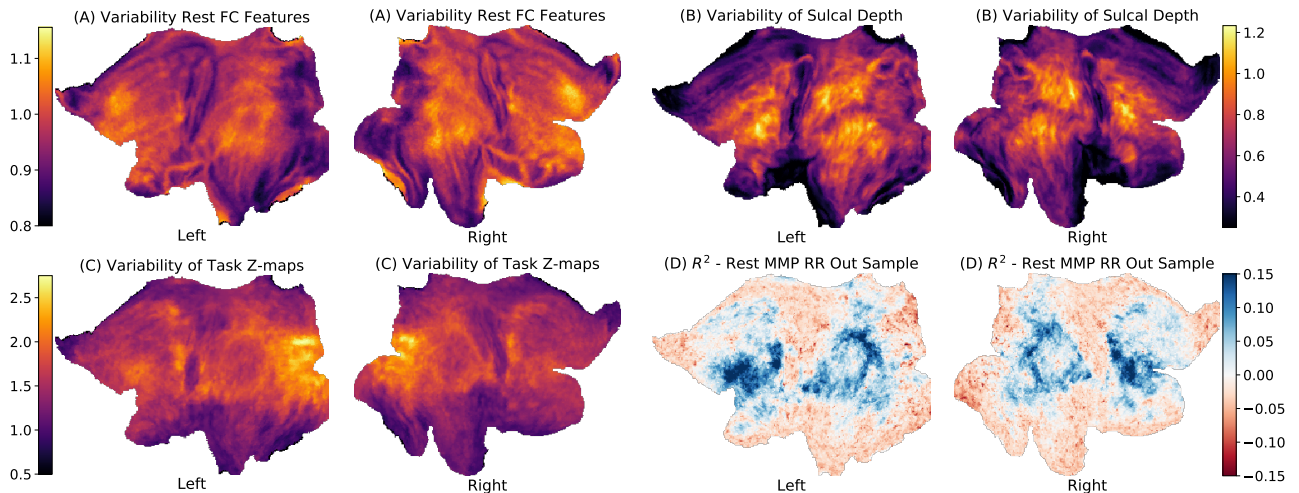
Figure 2.18: Flatmap projections of vertex-wise inter-subject variability as shown in figure 2.17. Shown is (A) **MMP-RR-PCR** FC Features, (B) sulcal depth, (C) task activation Z-maps. Plot D shows predictive R^2 of model **MMP-RR-PCR**.
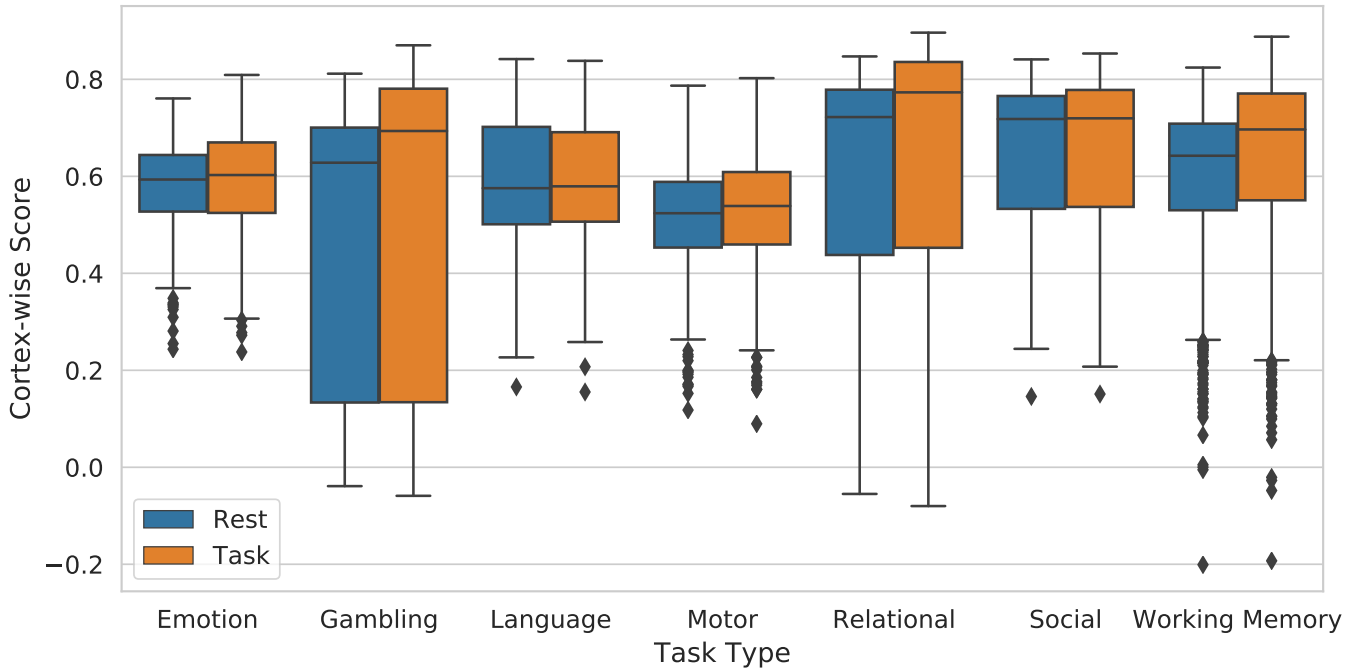
Figure 2.19: We investigated whether task data from separate task measurements provided better features for task prediction than resting-state data. Implicitly it is assumed resting-state scans, as opposed to task-based data where a very limited number of cognitive brain activity modes are investigated, subjects will enter multiple cognitive modes comprising of default, visual, motor, executive control and attention processes. This is supported by the networks of brain activity that are elicited during a single measurement of rest largely overlapping those that are extracted during the task. Furthermore, even across different task states, it appears that a core cognitive network dominates (Krienen et al. 2014). Do these multiple cognitive modes that are speculatively elicited during a rest scan differentiate subjects better than tfMRI?. Here, we tested whether a concatenation of data from the HCP battery of 6 diverse, but ultimately limited tasks. In the tfMRI case, separate features were calculated by selecting only 6 of the 7 tfMRI datasets, leaving out the tfMRI measurement of the to-be predicted GLM task contrast. Doing this excluded circularity. In reported experiments, data matrix X_i was a concatenation across 6 of the 7 task measurements with the excluded measurement being the one under which the contrast map was computed with. This led to features being computed from 3468, 3314, 3188, 3252, 3356, 3272, 3010 samples for EMOTION, GAMBLING, LANGUAGE, MOTOR, RELATIONAL, SOCIAL, WM contrast map predictions, respectively.)

2018). Even after application of spatial normalization transformations, considerable anatomical inter-subject variability is preserved despite liberal smoothing application. Additionally, echoplanar imaging (EPI) distortions due to B0 inhomogeneities and other individual specific factors, e.g., coil loading or other RF scaling issues, physiological, motion contaminants, and dependence of individual vascular factors to cortical orientation to B0 would reveal intra-subject dependencies between a rsfMRI and tfMRI acquisition. Regarding the dependence of individual vasculature, large signal biases on BOLD amplitude due to cortical orientations was shown to exist for 3T HCP data (Viessmann et al. 2019). This observation would undoubtedly create additional intra-subject dependencies between measurements that remain after normalization irrespective of any functional organization structure due to underlying neurophysiology or patterns of cortical computations. Therefore, a large degree of dependence will remain after applying normalization transformations and will not necessarily imply that intra-subject prediction scores are necessarily meaningful alone. Disentangling those factors remains to be explored in detail for future work. Second, the overall test-retest reliability of tfMRI is poor making individual difference research for fMRI difficult with most common task paradigms, especially considering the very limited number of task trials GLMs were computed over for HCP data (Elliott et al. 2019). We would therefore like to emphasize that considerable noise is present in estimates of first-level task effects we sought to predict. In this examination, no model considerations of it was incorporated into any design or analysis decisions. Last, activity summarized by a task GLM model is a useful measure only insofar as our a priori beliefs about how the task should be parameterized. Encoding models of the task that do not rely on strong assumptions of BOLD response may provide more powerful ways to summarize the kind of dependence we wished to characterize and remains an exciting avenue to explore beyond GLM maps (Lohmann et al. 2016).

2.7 Summary

Our closer examination using Human Connectome Project (HCP) data reveals that a majority of published models evaluated within our benchmark under current methods with many contrast targets examined did not perform better than naive, baseline models when only rsfMRI features and whole-cortex prediction were considered. This paper aims to remedy this issue and make a convincing case for utilizing methods to describe individual factors beyond merely remarking on individual differences. We propose single-vertex fitted methods that achieve a significant performance boost above baseline performance on the majority of contrast targets. Additionally, we provide benchmarks of comparable methods in published literature and include a variety of models with feature properties worth investigating, table

2.1. We provide further empirical characterization of top performing methods by an examination of showing where predictions performed well spatially. This explains why predictions of only a modest number of contrasts is possible above a naive baseline. Ultimately, we show that a model's prediction score can be taken as a relative measure of dependency between rest and task. These predictions results show a compelling behavioral correspondence to a subject's task performance committed during a tfMRI acquisition albeit with notable caveats. We hope that further improvements to this methodology will enable better understanding of rest-task correspondence informing individual behavioral measures.

2.7.1 Score Tables

Contrast	MMP-RR-PCR	Rest-Task GICA-RR	Rest-Rest GICA-RR	MMP-RR-DR	MMP-RR	AF-Mod	GRP-RR	Group Mean	RR-Anatomical	MMP-ParcelRR	GICA-DR-OLS	Group Z-Stat	MMP-OLS	AF
MEAN	0.582	0.581	0.58	0.578	0.574	0.571	0.568	0.561	0.557	0.551	0.55	0.54	0.409	0.302
REL	0.766	0.763	0.764	0.759	0.759	0.741	0.748	0.726	0.727	0.745	0.74	0.693	0.639	0.46
TOM	0.765	0.764	0.763	0.76	0.758	0.737	0.747	0.721	0.72	0.747	0.731	0.685	0.634	0.421
MATCH	0.751	0.749	0.749	0.745	0.745	0.731	0.734	0.713	0.714	0.736	0.726	0.679	0.621	0.479
MATH-STORY	0.729	0.739	0.735	0.73	0.732	0.729	0.716	0.695	0.694	0.737	0.705	0.642	0.609	0.618
2BK	0.739	0.735	0.735	0.73	0.729	0.709	0.714	0.713	0.69	0.701	0.705	0.666	0.604	0.397
RANDOM	0.733	0.731	0.73	0.729	0.725	0.708	0.718	0.692	0.69	0.708	0.695	0.654	0.585	0.433
PLACE	0.735	0.732	0.731	0.728	0.727	0.717	0.718	0.701	0.7	0.692	0.698	0.65	0.594	0.4
0BK	0.723	0.718	0.718	0.715	0.713	0.704	0.704	0.69	0.689	0.677	0.687	0.64	0.577	0.375
TOOL	0.712	0.708	0.708	0.704	0.701	0.691	0.691	0.678	0.676	0.68	0.681	0.644	0.562	0.399
0BK_PLACE	0.692	0.688	0.687	0.686	0.684	0.682	0.678	0.667	0.666	0.651	0.658	0.624	0.53	0.391
2BK_PLACE	0.693	0.691	0.69	0.687	0.685	0.674	0.676	0.66	0.658	0.658	0.661	0.628	0.539	0.368
REWARD	0.687	0.685	0.684	0.681	0.677	0.665	0.667	0.647	0.645	0.66	0.662	0.625	0.544	0.412
BODY	0.683	0.679	0.679	0.676	0.672	0.661	0.664	0.65	0.649	0.649	0.652	0.627	0.525	0.358
PUNISH	0.669	0.666	0.663	0.662	0.658	0.646	0.647	0.629	0.627	0.643	0.643	0.613	0.517	0.395
FACE	0.676	0.67	0.67	0.667	0.664	0.649	0.654	0.639	0.638	0.624	0.633	0.608	0.518	0.304
2BK_BODY	0.662	0.658	0.658	0.655	0.652	0.64	0.643	0.629	0.627	0.636	0.637	0.618	0.502	0.363
0BK_TOOL	0.653	0.65	0.65	0.647	0.642	0.643	0.636	0.631	0.628	0.627	0.627	0.602	0.476	0.367
CUE	0.643	0.651	0.649	0.645	0.642	0.639	0.635	0.62	0.619	0.632	0.612	0.572	0.478	0.421
CUE-AVG	0.644	0.651	0.649	0.646	0.642	0.638	0.636	0.621	0.621	0.633	0.617	0.552	0.489	0.405
2BK_TOOL	0.65	0.647	0.646	0.643	0.639	0.629	0.629	0.619	0.616	0.624	0.625	0.604	0.482	0.362
2BK_FACE	0.654	0.65	0.649	0.647	0.643	0.63	0.632	0.618	0.616	0.62	0.621	0.607	0.49	0.338
FACES	0.619	0.619	0.618	0.615	0.612	0.616	0.609	0.601	0.597	0.603	0.592	0.57	0.435	0.397
SHAPES	0.597	0.601	0.6	0.596	0.592	0.599	0.589	0.58	0.576	0.561	0.581	0.559	0.405	0.417
0BK_BODY	0.586	0.582	0.583	0.58	0.576	0.575	0.572	0.568	0.564	0.558	0.561	0.552	0.396	0.304
0BK_FACE	0.575	0.57	0.57	0.568	0.564	0.56	0.56	0.556	0.553	0.522	0.537	0.528	0.386	0.233
STORY	0.549	0.556	0.554	0.552	0.551	0.521	0.539	0.514	0.511	0.551	0.532	0.511	0.38	0.186
T	0.534	0.531	0.532	0.532	0.526	0.532	0.527	0.529	0.526	0.484	0.502	0.513	0.327	0.243
FACES-SHAPES	0.545	0.539	0.536	0.536	0.528	0.523	0.523	0.516	0.515	0.52	0.492	0.474	0.366	0.221
2BK-0BK	0.512	0.512	0.508	0.507	0.505	0.513	0.495	0.484	0.479	0.528	0.511	0.487	0.327	0.402
AVG	0.53	0.527	0.527	0.526	0.519	0.521	0.517	0.51	0.508	0.491	0.494	0.493	0.341	0.262
T-AVG	0.531	0.529	0.53	0.53	0.523	0.533	0.525	0.533	0.528	0.454	0.497	0.498	0.309	0.197
RF	0.52	0.517	0.516	0.515	0.508	0.514	0.507	0.508	0.505	0.472	0.48	0.49	0.327	0.201
LH	0.516	0.516	0.515	0.518	0.504	0.512	0.506	0.507	0.504	0.457	0.482	0.484	0.317	0.213
LH-AVG	0.517	0.519	0.518	0.518	0.51	0.516	0.51	0.52	0.516	0.453	0.487	0.503	0.3	0.09
RANDOM-TOM	0.496	0.492	0.49	0.489	0.484	0.497	0.479	0.477	0.474	0.49	0.466	0.485	0.301	0.353
LF	0.513	0.51	0.509	0.509	0.502	0.505	0.502	0.502	0.498	0.468	0.471	0.485	0.313	0.185
PLACE-AVG	0.504	0.497	0.496	0.496	0.493	0.507	0.489	0.493	0.487	0.471	0.443	0.489	0.296	0.281
MATH	0.504	0.501	0.501	0.501	0.495	0.473	0.484	0.474	0.471	0.501	0.488	0.479	0.327	0.122
RH	0.495	0.495	0.495	0.494	0.486	0.494	0.487	0.486	0.482	0.444	0.463	0.465	0.294	0.206
FACE-AVG	0.495	0.487	0.486	0.485	0.481	0.481	0.476	0.476	0.469	0.473	0.424	0.469	0.282	0.291
RH-AVG	0.486	0.487	0.487	0.486	0.48	0.487	0.481	0.492	0.486	0.419	0.458	0.469	0.267	0.115
RF-AVG	0.466	0.467	0.466	0.465	0.46	0.468	0.462	0.474	0.468	0.403	0.415	0.448	0.248	0.092
LF-AVG	0.465	0.463	0.462	0.463	0.457	0.466	0.458	0.472	0.466	0.4	0.413	0.435	0.245	0.055
BODY-AVG	0.398	0.394	0.395	0.394	0.388	0.392	0.388	0.398	0.392	0.369	0.356	0.404	0.196	0.108
MATCH-REL	0.353	0.354	0.354	0.356	0.346	0.363	0.348	0.35	0.341	0.333	0.336	0.364	0.168	0.269
TOOL-AVG	0.311	0.305	0.304	0.304	0.295	0.3	0.291	0.315	0.303	0.294	0.281	0.336	0.131	0.176
PUNISH-REWARD	0.096	0.096	0.096	0.095	0.093	0.102	0.093	0.118	0.105	0.086	0.091	0.173	0.026	0.086

Table 2.2: Correlation Score Table: Mean Correlation Scores averaged over subjects and shown across all models.

Contrast	MMP-RR-PCR	Rest-Task GICA-RR	Rest-Rest GICA-RR	MMP-RR-DR	MMP-RR	AF-Mod	GRP-RR	Group Mean	RR-Anatomical	MMP-ParcelRR	GICA-DR-OLS	Group Z-Stat	MMP-OLS	AF
MEAN	0.025	0.022	0.02	0.018	0.011	0.001	-0.002	-0.02	-0.098	-0.029	-0.076	-2.649	-0.686	-940632.771
REL	0.095	0.086	0.088	0.08	0.078	0.022	0.04	-0.021	-0.087	0.03	-0.062	-3.204	-0.537	-2297521.704
TOM	0.107	0.102	0.096	0.094	0.083	0.024	0.048	-0.022	-0.09	0.045	-0.056	-3.258	-0.511	-406035.33
MATCH	0.081	0.071	0.073	0.067	0.059	0.025	0.028	-0.022	-0.095	0.039	-0.066	-3.48	-0.596	-2716221.733
MATH-STORY	0.151	0.179	0.168	0.155	0.166	0.158	0.124	-0.019	-0.085	0.157	0.061	-1.963	-0.336	-282755.818
2BK	0.092	0.083	0.083	0.072	0.074	0.015	0.035	-0.024	-0.095	-0.014	-0.076	-1.045	-0.497	-1182910.045
RANDOM	0.081	0.077	0.074	0.075	0.065	0.019	0.033	-0.02	-0.093	0.02	-0.073	-2.771	-0.545	-4061690.458
PLACE	0.07	0.06	0.057	0.051	0.05	0.019	0.022	-0.022	-0.094	-0.052	-0.103	-1.867	-0.589	-965346.968
0BK	0.06	0.048	0.048	0.041	0.038	0.011	0.012	-0.023	-0.094	-0.063	-0.094	-1.192	-0.583	-1250291.134
TOOL	0.062	0.054	0.053	0.045	0.043	0.011	0.014	-0.021	-0.1	-0.023	-0.079	-1.494	-0.581	-1047548.118
0BK_PLACE	0.038	0.026	0.024	0.022	0.018	0.012	0.002	-0.021	-0.093	-0.063	-0.094	-1.975	-0.664	-546003.116
2BK_PLACE	0.052	0.047	0.046	0.039	0.035	0.009	0.012	-0.023	-0.101	-0.032	-0.091	-2.09	-0.627	-1033948.366
REWARD	0.067	0.061	0.057	0.057	0.049	0.019	0.023	-0.02	-0.094	0.007	-0.048	-3.146	-0.539	-976646.385
BODY	0.047	0.039	0.04	0.033	0.028	0.0	0.007	-0.023	-0.094	-0.033	-0.08	-1.613	-0.593	-1429015.359
PUNISH	0.06	0.055	0.048	0.049	0.042	0.014	0.016	-0.021	-0.098	0.006	-0.056	-3.31	-0.588	-666174.275
FACE	0.05	0.038	0.04	0.033	0.03	-0.007	0.007	-0.027	-0.102	-0.062	-0.114	-1.384	-0.595	-732123.083
2BK_BODY	0.046	0.04	0.039	0.034	0.032	0.001	0.011	-0.021	-0.09	-0.014	-0.056	-2.221	-0.562	-877546.337
0BK_TOOL	0.024	0.016	0.017	0.012	0.005	0.001	-0.011	-0.022	-0.113	-0.035	-0.077	-1.87	-0.738	-825311.656
CUE	0.027	0.044	0.041	0.035	0.02	0.019	0.002	-0.017	-0.089	0.014	-0.098	-9.776	-0.688	-5853287.768
CUE-AVG	0.025	0.038	0.035	0.031	0.011	0.012	-0.005	-0.019	-0.096	0.013	-0.097	-11.297	-0.682	-5293763.983
2BK_TOOL	0.039	0.033	0.032	0.027	0.023	-0.001	0.002	-0.021	-0.1	-0.016	-0.065	-1.713	-0.621	-890898.033
2BK_FACE	0.046	0.038	0.038	0.033	0.03	-0.001	0.006	-0.024	-0.104	-0.023	-0.086	-1.661	-0.61	-480804.67
FACES	0.014	0.015	0.012	0.008	0.001	0.009	-0.008	-0.019	-0.101	-0.013	-0.069	-3.65	-0.746	-880524.147
SHAPES	0.008	0.014	0.013	0.008	-0.004	0.012	-0.01	-0.019	-0.098	-0.044	-0.046	-3.64	-0.734	-538566.554
0BK_BODY	0.006	-0.001	0.0	-0.002	-0.011	-0.011	-0.019	-0.022	-0.104	-0.043	-0.073	-1.535	-0.732	-1245430.971
0BK_FACE	0.005	-0.004	-0.003	-0.006	-0.011	-0.017	-0.018	-0.025	-0.098	-0.077	-0.1	-1.599	-0.711	-617639.447
STORY	0.026	0.033	0.031	0.03	0.028	-0.009	0.013	-0.016	-0.075	0.025	-0.019	-4.435	-0.536	-482269.634
T	-0.016	-0.018	-0.017	-0.018	-0.025	-0.018	-0.025	-0.02	-0.113	-0.085	-0.092	-1.383	-0.898	-236227.856
FACES-SHAPES	0.017	0.011	0.004	0.006	-0.003	-0.013	-0.011	-0.02	-0.09	-0.018	-0.099	-2.615	-0.659	-247146.447
2BK-0BK	0.017	0.01	0.008	0.008	0.002	0.015	-0.01	-0.019	-0.105	0.032	-0.019	-1.815	-0.668	-10010.264
AVG	0.011	0.006	0.007	0.007	-0.003	-0.001	-0.006	-0.016	-0.092	-0.037	-0.076	-1.774	-0.718	-194630.22
T-AVG	-0.028	-0.03	-0.029	-0.029	-0.039	-0.026	-0.036	-0.023	-0.121	-0.139	-0.104	-1.13	-0.984	-86920.081
RF	-0.002	-0.005	-0.006	-0.005	-0.016	-0.01	-0.016	-0.017	-0.091	-0.059	-0.093	-2.119	-0.728	-179995.523
LH	-0.005	-0.006	-0.007	-0.005	-0.019	-0.011	-0.018	-0.016	-0.092	-0.074	-0.078	-2.346	-0.709	-63500.533
LH-AVG	-0.022	-0.02	-0.023	-0.021	-0.032	-0.024	-0.034	-0.018	-0.091	-0.102	-0.092	-3.225	-0.826	-49450.47
RANDOM-TOM	0.004	-0.001	-0.001	-0.002	-0.009	0.007	-0.015	-0.015	-0.084	-0.004	-0.055	-4.323	-0.683	-18761.948
LF	-0.003	-0.005	-0.007	-0.005	-0.015	-0.013	-0.016	-0.017	-0.092	-0.058	-0.093	-2.071	-0.769	-331533.112
PLACE-AVG	-0.007	-0.017	-0.017	-0.016	-0.02	-0.002	-0.025	-0.02	-0.113	-0.048	-0.133	-1.846	-0.805	-15150.453
MATH	0.013	0.009	0.009	0.011	0.005	-0.02	-0.009	-0.018	-0.083	0.01	-0.035	-4.228	-0.593	-242997.845
RH	-0.006	-0.008	-0.007	-0.008	-0.018	-0.009	-0.017	-0.017	-0.091	-0.062	-0.075	-1.98	-0.763	-49992.747
FACE-AVG	-0.002	-0.013	-0.014	-0.015	-0.02	-0.019	-0.025	-0.023	-0.112	-0.028	-0.116	-1.482	-0.818	-210027.135
RH-AVG	-0.026	-0.025	-0.025	-0.026	-0.033	-0.025	-0.032	-0.02	-0.101	-0.096	-0.081	-2.147	-0.828	-44493.434
RF-AVG	-0.027	-0.025	-0.026	-0.027	-0.033	-0.024	-0.031	-0.017	-0.096	-0.096	-0.12	-2.259	-0.819	-27104.002
LF-AVG	-0.026	-0.028	-0.03	-0.028	-0.034	-0.025	-0.033	-0.017	-0.104	-0.096	-0.123	-2.557	-0.858	-217395.485
BODY-AVG	-0.022	-0.026	-0.025	-0.025	-0.033	-0.026	-0.032	-0.02	-0.109	-0.043	-0.096	-1.575	-0.853	-91476.217
MATCH-REL	-0.02	-0.017	-0.017	-0.015	-0.022	-0.008	-0.018	-0.021	-0.099	-0.032	-0.051	-3.646	-0.752	-20170.165
TOOL-AVG	-0.026	-0.032	-0.032	-0.031	-0.04	-0.034	-0.042	-0.021	-0.111	-0.033	-0.062	-1.173	-0.836	-20230.362
PUNISH-REWARD	-0.036	-0.036	-0.035	-0.036	-0.038	-0.03	-0.037	-0.02	-0.114	-0.034	-0.047	-1.622	-0.949	-75151.823

Table 2.3: Weighted R^2 Score Table: R^2 Scores across all models examined ordered by Contrast names.

48

Chapter 3

From Connectome-fingerprinting to Learning-fingerprinting

Although mapping rsfMRI features to tfMRI GLM maps may be interesting in its own right due to its practical relevance detailed earlier, tfMRI GLM maps ultimately seek to explain behavior and cognitive function. They largely do so by localizing where brain activity correlates with isolated processes representative of some feature believed to be involved in performing the task. Behavioral investigations related to connectome fingerprinting so far, and as shown in the previous chapter, have focused on whole-brain summaries describing overall, whole-brain statistical dependence. Akin to most investigations of spontaneous activity's predictive ability to identify psychological traits and clinical diagnoses, the previous chapter provided evidence that dependence between rsfMRI and tfMRI correlates with behavioral performance measured during the task. These type of predictions are formed via machine learning models that resemble "black-boxes", relying on what we can broadly call connectome biomarkers. Statistical summaries of whole-brain biomarkers or rest-task dependence, however, do not tell us anything about what specifically may underlie why such mappings should be successful at all. Progress on that front demands an investigation as to what cognitive and physiological process may be responsible for driving such correspondence. Doing so requires the design of a hypothesis that carefully isolates those processes to test with valid methodology.

Given the desire to understand motor skill learning in our context, I seek to detail the design of a hypothesis to investigate that in detail. In particular, asking whether methods of connectome fingerprinting detailed earlier have an ability to say anything meaningful about behavior beyond abstract statistical summaries measuring dependence. Instead, I ask, can

these methods be used to spatially map learning-related effects, i.e., a learning fingerprint. Further, can a learning fingerprint be used to examine learning-related effects of a motor learning paradigm that has so far attracted little or no attention in human neuroimaging investigations. Following this line of questioning, the chapter details with a heavy behavioral focus:

1. What physiological process we speculate might produce the differential effects caused by learning we see with fMRI.

2. What we know about differential learning-related effects seen in both rsfMRI and tfMRI separately and together in longitudinal investigations so far.

3. How learning is measured with current behavioral paradigms in sensorimotor skill learning.

4. How utilizing a new motor learning paradigm investigating plasticity-related effects covers ground not performed in most neuroimaging investigations.

3.1   Physiological Underpinnings

Critically, if a restructuring of neural activity occurs, e.g., through experience or learning, then the restructuring should be reflected in both spontaneous and task-evoked activity. On a more concrete level then, what physiological process, i.e., mechanism, might govern the restructuring observed in brain activity? In answering that, we will be capable of identifying more precisely why task activity should relate to spontaneous activity in a behavioral context.

Broadly stated, we understand neurons function in large, complex interconnected networks. These networks rely on signals to communicate between populations of neurons that implement local computations scattered across the entire cortex. The reorganization of how these neurons coordinate to act together are understood to be the the basis of learned, new or adapted behavior. This evolves under neural plasticity: the adding, removing, strengthening, and weakening of neuronal elements functioning in these networks. Neural activity, whether internal or externally generated, is the primary driver for these adapative changes in the CNS. Changes into new functional states of the CNS are observed on the level of synapses, neurons, neural circuits, populations, and macro-scale whole-brain networks (Buonomano & Merzenich 1998; Mattar et al. 2018). They have been detected in the underlying structure of brain anatomy and within the functional properties of neurons.

On a synaptic level, Hebbian learning, is the canonical thesis explaining how neurons alter

their coordination to support learning (Hebb 1949). Since Hebbian learning was originally proposed over five decades ago, considerable progress has been made at understanding how neural activity underlying behavior modifies the function and structure of neural circuits (Karmarkar & Dan 2006; Ganguly & Poo 2013). For instance, on a microscale, physiological processes of long-term potentiation (LTP) (Bliss & Lømo 1973) and long-term depression (LTD) (Ito & Kano 1982) explain a mechanism responsible for the modification of synapses. Other mechanisms that have been proposed include Spike-Timing-Dependent Plasticity (STDP) (Gerstner et al. 1996; Markram et al. 1997), or more recently, Differential Extrinsic Plasticity (Der & Martius 2015).

These synaptic level plasticity mechanisms are understood as the theoretical basis explaining the development of new or altered neural circuits. Currently, we understand activity dependent plasticity either alters tuning properties of individual neurons or, as more plausible recent evidence highlights, change how populations of neurons interact between each other (Seriès et al. 2004). This kind of population tuning is hypothetically understood to take place in 3 sequential steps (Makino et al. 2016):

1. Exploration: behavioral variability generates an increase in the number of neural populations recruited. The exploration phase of the hypothesized population tuning mechanism may be driven by a reduction of gamma-aminobutyric acid (GABA) (Floyer-Lea et al. 2006). This generates a reduction of inhibitory activity creating an increase in neuronal variability.

2. Selection: a solution is found that is capable of satisfactorily accomplishing the task and a strategy of movement selected.

3. Refinement: the solution can be exploited to lead to better speed and accuracy, improved energy economy, and a reduction of movement variability. These changes involving the formation of new circuits provide the means for new behavior responsible for learning phenomena (Oby et al. 2019).

3.2   Evidence for tfMRI and rsfMRI learning changes

What evidence exists so far suggesting functional changes incurred by learning is detectable with tfMRI and rsfMRI? To date, fMRI remains probably the best non-invasive human modality to provide insights into learning-dependent plasticity at the scale of large, whole-brain functional networks. However, investigations examining direct physiological changes with invasive techniques and fMRI simultaneously that accompany learning associated ef-

fects remain few and far between (Álvarez-Salvado et al. 2014). Although we can speculate that the physiologcial processes are Hebbian in nature, we at present have little converging evidence lines to clearly support that. Therefore, mostly differential changes under the indirect neural measures of fMRI allow one to make any inferences (Poldrack 2000). Depending on the measurement and analysis paradigms used, these differential changes can largely be divided between increases and decreases in activation magnitude, changes in correlation strength, or more generally, changes in cortical representation. Longitudinal studies make up the vast majority of learning studies, where subjects are measured repeatedly multiple times over a period of learning session. The interpretation of these studies are finely nuanced and comprise of many potential neural or cognitive effects among a host of potential confounds on both a cognitive and image acquisition level (See (Poldrack 2000) for a comprehensive review).

3.2.1 OBSERVING RSFMRI PLASTICITY CHANGES

The use of rsfMRI for examining learning-related effects are broad and many. Limitations due to the scanner environment can give considerable difficulties in measuring and performing complex tasks for extended periods. Experimental designs measuring learning-induced changes with resting-state acquisitions allow subjects to train outside of the scanner environment remedy this issue. Many have met considerable success. Studies examining adult human plasticity can roughly be catagorized between (1) the specific learning modality: sensory/perceptual vs. motor vs. cognitive learning, and (2) the length of time under which learning effects are examined. Generally speaking, rsfMRI signal correlations are thought to reflect prior history of co-activations between brain areas (Guerra-Carrillo et al. 2014). Investigations motivated by this belief form the basis of most rsfMRI plasticity change hypotheses.

Albert et al. appears to have published the first known study relying on rsfMRI measurement examining a sensory-motor adaptation task within a single, short-term measurement session (Albert et al. 2009). Comparisons between a naive pre-trained state and post-trained state showed increased cerebellar and frontoparital involvement in the learned group over a control group of subjects. Longer measurement periods over the course of days or months have also been conducted where processes related to early and late learning have been identified. For instance, Lewis et al. show that training effects manifest after multiple days of intense training on shape identification (Lewis & Baldassarre 2009). Interactions between visual cortex and frontal-parietal areas appear to reconfigure and these changes appear correlated with the degree of perceptual learning.

52

Similar work has been conducted in many other domains. Groups have examined force-field adaptation (Sidarta et al. 2016), sequence learning (Bonzano et al. 2015), perceptual learning (Guidotti et al. 2015), cognitive reasoning (Hearne et al. 2017), challenging balance tasks (Taubert et al. 2011), musical training tasks (Amad et al. 2016; Wollman et al. 2018), and motor learning by observation (McGregor & Gribble 2017). The most compelling recent study is by Newbold, et al. where they show plasticity related effects relying on a dense-sampling approach (Newbold et al. 2020). Three volunteers had their dominant arm cast for 2 weeks such that dramatic disuse and weakness occurred. Changes were observed over the course of daily, 30 minute rsfMRI measurements. Analysis of these measurements revealed brain networks responsible for generating movement of their arm functionally "disconnecting" from the rest of their motor system. Interestingly, spontaneous activity in the form of large BOLD pulses appeared when disuse occurred. It is speculated that these pulses are a product of the brain attempting to maintain its old functional architecture needed for regular arm use.

3.2.2 OBSERVING TFMRI PLASTICITY CHANGES

Measuring task-evoked activity over resting-state holds the major caveat that performance of the task needs to usually take place within the scanner bore. This implicates that most task-based examinations of learning focus on either sequence (Telesford et al. 2017) or perception learning (Guidotti et al. 2015) tasks. Nevertheless, given fMRI's early focus on elucidating localized activity maps, tfMRI studies examing plasticity make up some of its earliest studies (Kami et al. 1995; Karni et al. 1998). These early studies demonstrate that recruitment of specific networks appear depending on either early or late stages of learning.

These studies together show that across the period of learning, often non-monotonic and complex changes in neural activity are observed (Wiestler & Diedrichsen 2013). Both activity increases due to increased recruitment of areas (Grafton et al. 1995a) (usually associated with early stages) and activity decreases (usually associated with later stages) due to accompanying changes of higher neural efficiency have been observed (Poldrack 2000). Due to that fact, they remain difficult to interpret in group studies, especially since these observations are not exclusive to one another and may occur simultaneously (Penhune & Steele 2012). This generally presents a challenge to interpreting these studies as averaged activity changes presents severe deficiencies due to the complex, non-monotonic changes noted.

3.2.3 rsfMRI and tfMRI Plasticity Changes Together

Studies showing plasticity changes incurred by learning in both rsfMRI and tfMRI in areas that remain congruent between detected differential changes remain understudied. Yet, they are critical; if we understand that plasticity changes occur due to a restructuring primarily driven by Hebbian-like task-evoked processes discussed earlier, then identifying areas elicited by the task is an essential a priori marker to investigate the differential changes detected within any modality, e.g., rsfMRI or structural. Also, differential changes across any MR modalities in longitudinal studies would appear to be very sensitive to false positives. This is especially the case in structural investigations relying on questionable methodology like VBM (Thomas et al. 2009). Therefore, in order to identify behavioral learning effects with the necessary specificity, they should be tackled via multiple modalities and careful longitudinal paradigms.

So far, three studies examining task-specific effects stand out in this regard. First, Harmelech, et al., show using a neurofeedback paradigm that activity increasingly evoked within the dorsal anterior cingulate cortex (dACC) produces corresponding activity changes observed in resting-state measurements within individual subject measurements (Harmelech et al. 2013). The intensity of functional correlations appearing at rest linearly relate to the level of prior activation measured during the neurofeedback driving recorded task-evoked activity. Second, Steel, et al., isolate training-related rsfMRI functional correlation changes specific to task-evoked responses in the form of offline considation (Steel et al. 2019). In order to do so, the study relied on two separate learning tasks compared to a non-training condition in order to establish specificity of effects measured over the course of multiple days. Additionally, effects related to time-of-day and behavioral performance of the task were incorporated in model design considerations giving a compelling demonstration of learning effects detectable within rsfMRI. Third, Shannon et al. use combined PET, fMRI, and DTI to examine the metabolic correlates of learning induced changes in a visuomotor adaptation task (Shannon et al. 2016; Magistretti 2016). Previous work from the same group suggested that a specific metabolic pathway of glucose—aerobic glycolysis—provided the necessary conditions to allow brain plasticity to occur. This idea mainly stemmed from PET data showing that early stages of human development have highest aerobic glycolysis levels throughout the brain (Raichle & Mintun 2006). Shannon et al. provide compelling support for effects of sensory learning in brain areas that are congruent across (1) evoked by the task, (2) show rsfMRI pre-training vs. post-training effects, and (3) follow that aerobic glycolysis as measured by PET increases in specific brain areas.

3.3 Examining Learning from Behavior

The previous section provides a background on why and how rsfMRI and tfMRI measurement paradigms could create a means to observe plasticity related effects over the course of learning. In particular, there was a focus on providing justification as to why a connectome fingerprinting approach might yield insights into mapping learning related effects. However, up until now, a discussion of behavioral factors has been absent. How should one understand what actually constitutes learning? Characterizing brain structure and activity alone without a strong theoretical and experimental focus of behavior places neuroscience research in a pallid position (Krakauer et al. 2017). Therefore, it is pertinent that research questions focus on behavior before neuroimaging investigations take place.

Here, we discuss the design of a novel task intended to utilize the aforementioned methodology of connectome-fingerprinting. To understand its novelty, a background context of established motor skill learning paradigms are briefly described. Many of these paradigms have already undergone a neuroimaging investigation and are discussed where appropriate.

3.3.1 WHY SENSORIMOTOR SKILL LEARNING?

How intelligent behavior is enabled by the brain and body within the domain of human motor control still remains a complete mystery. Task specific intelligent behavior, where the outcome is primarily driven by coordinated interaction with the environment, is referred to as **motor skill**. Vertebrate brains primarily evolved to support the development of skill in order to cope with the demands of a dynamic, competitive environment to accomplish biological goals (Llinás 2002). That is, an environment with low certainty, high risk, and swift changes in which to reproduce, fight for resources, promote feeding, among other vital behaviors. The ability to develop new skills to accomplish these goals is essential for continued survival. Thankfully, as humans, we have been on the receiving end of this evolution; our inherited biological endowment allows an extremely broad and diverse skill set to be achieved.

For instance, there are highly complex skills, e.g., performing an elaborate Chopin piece or playing table tennis, where the integration of many cognitive strategies, perceptual feats, and fine motor skills are needed to accomplish the task. These tasks may take many years of development with rigorous practice to accomplish well. And then there are simple skills, e.g., reaching an arm to a target area or performing a motor sequence of finger-thumb opposition movements that may require little or no conscious effort at all. These tasks after a period of early development appear to adapt very quickly to new environments and challenges.

Simple or complex, motor skill learning, or more generally speaking–sensorimotor learning–encompasses the entire perception-cognition-action loop. That is, the perceptual, associative, and motor execution processes needed to accomplish a motor task goal. This makes the study of human motor control and learning unique in neuroscience; their processes must integrate many different systems and levels of description–from fine neural substrates of physical mechanisms to abstract representations of task goals. Sensory systems and networks involved in perceptual learning alter areas involved in movement and areas involved in motor control change in response to areas involved in perception (Censor et al. 2012; Ostry & Gribble 2016). To appreciate this fact, with a rather arbitrary distinction, the development of a motor skill involves:

1. **Sensory perception** – processing information received from our senses, e.g., visual, proprioception. Abstract concept examples: visual search, pattern recognition, representation learning. Concrete example: estimating the position and orientation of a moving object such that relevant features are identified and exploited.

2. **Cognitive** – associations between information represented from the senses interact with memory, rewards, and mental concepts. Abstract concept examples: decision making, reinforcement, associative long-term memory, short-term memory, chunking. Concrete example: a verbal description of an action needs to be translated into a motor pattern.

3. **Motor** – interactions with the environment through refinement of executed movement. Abstract concept examples: motor programs, central pattern generators. Concrete example: basic rhythmic stepping to locomote smoothly.

Given the breadth of involved systems, there is no privileged area of the brain that we can generalize to all motor behaviors. The question of how sensorimotor skill learning takes place therefore needs to carefully consider how nearly every system of the brain are recruited to integrate together. For example, prefrontal and sensory associated areas relay neural activity to pre-motor and motor areas that plan and execute movements. This activity is generated by sensory systems that learn an efficient representation of sensory information, which relay to cognitive systems that select relevant strategies. Simultaneously, subcortical regions and specialized nuclei therein are strongly interconnected to areas known to have integral functions in movement generation. These regions, e.g., basal ganglia, are known to be important governing reinforcement processes or in selecting the appropriate motor pattern to be executed. Further, there are no learning mechanisms we can generalize to all components/systems involved, where each may be governed by separate learning related

processes. The next sections describes how motor skills are usually characterized and what paradigms are utilized to implement research goals. Together they strive to explain how an integration of this activity takes place to support learning.

3.3.1.1 Motor Skill Learning Paradigms - The Present Stage for fMRI

Motor skill is usually determined by the frequency (De Jong 1957) intensity of practice (Snoddy 1926). Typically, motor skills are learned to reach some asymptotic level where no further or little improvement is observed. Here, level is considered with respect to time and space: how fast the execution took place and how far the movement deviated from its location target. These quantities are not independent of each other and need to be optimized according to a speed-accuracy trade-off (Wickelgren 1977). However, speed and accuracy are an over-simplification and insufficiently explain skill alone; often movements do not optimize with respect to speed, but according to its energy economy, e.g., locomotion, or by a reduction of variability (Hasson et al. 2016). The later may mark different stages of learning; early stages involve exploration where high variability and task strategy switching may occur. After early stages with task strategy at hand, refinements under intense repetition achieve a highly stereotyped movement having low trial-to-trial variability (Makino et al. 2016).

Regardless of intermediate details, all motor skill development follows a canonical progression. The improvement process involves training over multiple sessions where skill acquisition takes place very quickly initially, i.e., fast learning, and then slows down considerably later on, i.e., slow learning (Karni et al. 1998), taking on the shape of an exponential curve. The parameters that govern this exponential curve will be specific to the task, e.g., learning a finger sequence versus learning a complex movement pattern requiring fine control. Hence, what defines a fast or slow stage of learning will depend on the task (Dayan & Cohen 2011). Different processes are known to govern the evolution of this curve; changes may occur during training (online) or improve after a training session finished (offline). Offline processes are known to occur under motor memory consolidation (Shadmehr & Holcomb 1997).

Extensive research into motor skill learning from decades prior has established some key concepts and paradigms. Fundamentally, motor skill learning is a question of *how* and *what* is learned to achieve a task goal. Asking how and what will heavily depend on the learning paradigm. A brief summary of these paradigms will aid in understanding the novelty of the approach to understand complex skill presented immediately followed in the next section.

Sensory Motor Adaptation. A ubiquitous feature of human motor control is the ability

to adapt end-effector trajectories under sensory perturbation. The motor system needs to accomplish that while accommodating noise and delay in sensory feedback, compensating for variability in movement generation. Known as sensory motor adaptation, paradigms used to study this phenomenon primarily include perturbations from loaded weights, prism glasses, and applied force fields (Von Helmholtz 1867; Shadmehr & Mussa-Ivaldi 1994; Wolpert et al. 2011; Shadmehr et al. 2010; Reichelt et al. 2013; Herzfeld et al. 2014; Buckingham et al. 2014). Under these paradigms, computational control models are the central argument to explain adaptation processes. Briefly explained, these models describe the existence of an **internal model**. An internal model encodes a control policy for an end-effector movement, i.e., a mapping of feedforward and feedback command (Sutton et al. 1998; Todorov & Jordan 2002). How these control policies evolve to adapt motor behavior is through sensory prediction errors. Errors are the difference between the actual, "true" sensory information received and an internal model prediction that creates the expected sensory consequence of the motor command. Perturbations toward sensory information, e.g., haptic or visual properties, provide the means to test how these internal models are formed, cued, or consolidated in the brain. Motor skill improvements reflect more efficient and accurate movement generation of limbs. Results from the use of sensory motor adaptation paradigms suggest that this is due to an internal model responsible for force and limb trajectories calculation, e.g., to solve the inverse problem (Lalazar & Vaadia 2008).

Typically, human studies have been tackled through executing a reaching movement to a target via a robotic manipulandum. There, perturbations are applied to a guided robot arm via force-fields, changing limb dynamics in a plethora of ways. Additionally, visuomotor transformations from altered visual feedback can be used. Most commonly, a visuomotor rotation or gain is imposed. Accuracy is measured as the deviation from the desired target. These studies, among many other findings, have yielded insights into the nature of feedback control (Todorov & Jordan 2002; Wolpert et al. 2011), impact of learning rates (Castro et al. 2014), consolidation (Krakauer et al. 2005), and Bayesian sensorimotor processes (Körding & Wolpert 2004). Together they make up the a rich literature describing algorithms involved in error-based (supervised) learning, reinforcement learning, and cognitive-level policies needed for motor control.

Neuroimaging investigations that involve the same or similar hardware as these behavioral investigations are limited. This is primarily because required robotic hardware remains difficult to bring into a scanner environment. Therefore, most evidence supporting the neural substrates of sensory motor adaptation stem from lesion studies (Smith & Shadmehr 2005; Shadmehr et al. 2010). Overall, the cerebellum has strongly been implicated in neural

investigations of adaptation (D'Angelo 2018). fMRI investigations largely reinforce this belief (Blakemore et al. 2001; Shadmehr & Holcomb 1997). Beyond the cerebellum, more recent investigation in mice have intriguingly found that the primary somatosensory cortex is also involved in updating internal models (Mathis et al. 2017).

Sequence Learning. In order to obtain a repertoire of complex movement for goal-oriented behavior, new sets of actions need to be assembled to work together (Lashley 1951). The skill of assembling sequences is an essential property of most everyday behavior. Paradigms that study the process of assembling temporally ordered movement are known as sequence learning. Sequences can be studied as a discrete set of actions, a continuous overlapping movement, or some hybrid in between (Krakauer et al. 2011).

The first type is by far the most prevalent where discrete sequences are generated in what is known as a Serial Reaction Time Tasks (SRTT) (Nissen & Bullemer 1987). SSRT paradigms evaluate the response of visual cues to generate ordered movements via what are usually finger button presses. The goal is to respond to the visual cues as fast as possible without anticipation for the sequence. Each target in the sequence appears with fixed delay until the required performance of the sequence is fully completed.

Overall, the ease SRTT tasks can be manipulated in a variety of environments, and the short amount of time for behavioral improvement to appear, make it an attractive paradigm to use, especially in neuroimaging where movement constraints are easily met. Reaction times and error rates assess improvement in a straightforward way. Further, trained sequences versus untrained sequences can be easily compared so control conditions within longitudinal studies of learning are possible. Animals can also be used for many comparative studies, too (Jin & Costa 2010).

The attractiveness of sequence learning paradigms has created a flood of neuroimaging investigations. These studies are broad, many, and without clear consensus given the literature's bias to report novel results (Berlot et al. 2020). For example, early findings suggested that activity increases (Grafton et al. 1995b; Kami et al. 1995; Hazeltine et al. 1997) and decreases (Jenkins et al. 1994; Toni et al. 1998; Ungerleider et al. 2002) were created by learning effects. A more recent, pre-registered, and high-powered investigation at ultra-high-field (UHF) MRI reveal parietal and premotor areas show decreases in overall activation (Berlot et al. 2020). Further, reorganization of sequence-specific patterns appear early in learning and outside of primary motor regions, e.g., M1.

Skill Learning - De novo. Acquisition of an entirely new control policy are often needed to

map movement to achieve task goals. De novo examinations ask how the generation of novel movement occur. This contrasts to paradigms, e.g., sequence and sensory motor adaptation, where a recalibration of an existing control policy takes place. Instead, de novo learning involves an acquisition of a new control policy requisite to learn some arbitrary association between a stimulus and action.

De novo learning remains the least investigated from all other paradigms, and because of the flexibility in examining arbitrary movements, probably the most diverse (Krakauer et al. 2011). The simplest behavioral studies involve mapping new, arbitrary visual associations to a set of discrete motor responses. For instance, recognizing a figure or shape presented and pressing the correct button in response. Learning is a measured by accuracy relative to response time, i.e., speed-accuracy trade-off. More sophisticated examinations focus on continuous skill development. For instance, an arbitrary mapping between hand posture and motion might map to an auditory or visual representation. Mirror reversal paradigms are another popular example requiring a new control policy that make an easy comparison/contrast with adaptation paradigms (Yang et al. 2020).

Behavioral and neuroimaging investigations so far remain sparse (Choi et al. 2019). Different neuronal mechanisms are thought to be responsible for this process compared to sequence or adaptation learning. It is unclear what kind of overlap exists between these paradigms. More broadly, however, because of the formative need for action selection, feedback like reward appears to play a central role. Therefore, the basal ganglia has strongly been tied to this kind of motor learning. Further, separate mechanisms are involved from early goal-directed behavior needed to explore correct mappings to later more automatic behavior needed to refine the motor output.

3.3.1.2 Towards a more embodied paradigm

The motor skill learning paradigms discussed so far emphasize representations largely centered in the brain. Further they focus mostly on tasks that involve very simple movement in simple interactions with its environment. These tasks can be easily isolated in lab environments, e.g., reaching, pointing, or generating finger sequences. While this allows highly controlled manipulations, they lack in other aspects that no neuroimaging investigation has so far focused on. I discuss why a different approach might yield considerable insights given its emphasis on other control aspects that are largely not found in physiological studies.

Biological or synthetic, any system's behavior is not *solely* governed by the CNS or some "top-down" control mechanism. Behavior is also the product of properties unique to the
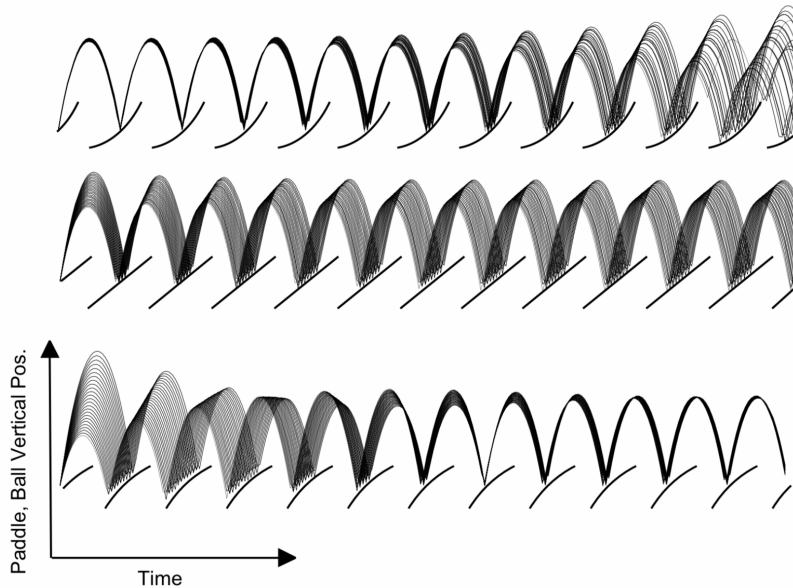
Figure 3.1: A simulation of the ball trajectory from an open-loop system having a simple sinusoidal paddle trajectory. Shown are three different trajectories with different paddle impact acceleration. From top to bottom, paddle impact acceleration is positive, zero, and negative. In the negative regime, the ball trajectory converges to a stable fixed point while in the positive regime, the ball diverges and errors increase. Adapted from (Schaal & Atkeson 1993)

body and environment it is situated in (Brooks 1990). Some salient examples include shape of its body morphology, corresponding material properties of its morphology, or particular influences of its ecological niche where motor tasks are situated (Pfeifer et al. 2007). These examples are well-developed in the wide-reaching literature of *embodidment* and dynamical systems research. Broadly summarized, physical, environmental, and task constraints afford the ability of a dynamical system to self-organize and complex behavior to emerge. In the case of the CNS, its role centers on exploiting these organizational patterns such that any needed (and potentially costly) neural information processing can be explored, refined, and executed (Lungarella & Sporns 2006). An embodied agent, therefore, has its sensors, limbs, actuators, and internal control units, i.e., nervous system, highly coupled to one another and its environment; one small change in any component and information flow between any of these components and behavior can be dramatically altered non-linearly. This perspective emphasizes that behavior shapes cognition as a coupling between physical and information theoretic processes. Therefore, behavior cannot be understood strictly as a control process centered from some top-down representation.

Why is this perspective useful and needed? Take Bernstein's degrees of freedom or motor equivalence problem as one example (Bernstein 1967). The human body's joints, limbs and muscles overspecify the number of variables needed for a solution to the motor control task.

Given the enormous repertoire of movements available to accomplish a single motor task, it quickly becomes highly implausible that the CNS recruits a separate encoding of each motor solution it takes. Redundancy defined by overspecified muscles and joints, overspecified kinematics, and finally overspecified neural circuits exist. No compelling evidence supports the idea that this amount of modularization is present in neurophysiology (Georgopoulos et al. 1982). Similarly, motor tasks themselves may have a considerable degree of overspecification/redundancy where different solutions to task variables have the same outcome (Müller & Sternad 2009). Therefore, simpler control algorithms must be present, but the question remains how, where, and why they emerge. By understanding that motor control emerges from the coordination of physical, environmental, and task constraints, we can gain better insight into those algorithms. However, to enable a rigor that can quantify human behavior under this perspective, one severe problem remains: the curse of dimensionality. In the real world, it quickly becomes untenable to specify the degrees of freedom the human brain, body, and environment have together. Therefore, a considerable reduction in available degrees of freedom, especially of the motor task, is necessary to isolate motor control needed to study dynamical properties of the task.

The simplest starting point is to focus on tasks that can have their dynamics completely specified mathematically. One such system is virtual paddle juggling where an agent rhythmically bounces a ball to a specified height. The task can be simplified to an interaction in one DOF. With this simplification, the system is entirely governed by periodic motion of the paddle, gravitational forces, and the exchange of energy to the ball during instantaneous impact. Given this reduction in complexity, the task has evolved to become a model system for investigating the behavioral dynamics of rhythmic movement (Siegler et al. 2013).

Despite its simplicity, the task involves a combination of active and passive control strategies to accomplish (Ronsse & Sternad 2010). On passive control, the task dynamics have identifiable analytic properties that, when found, offer clear control advantages. These control advantages have been examined in human behavioral studies; participants performing rhythmic bouncing show that they exploit dynamics of the ball-paddle interaction such that the task can, to some degree, be "self-controlling." In the original seminal behavioral study of paddle juggling (Schaal et al. 1996), Schaal et al. demonstrate an analysis of this property; under rhythmic performance of the task where the periodic trajectory of the racket is stable, participants learn to interact with the ball with negative racket acceleration. In this negative acceleration regime, if the racket movement continues in an open-loop configuration, then perturbations that would disrupt the movement will diminish and the ball will converge to its nominal height. A simulation showing this feature is available in figure 3.1. These

observations suggest that the acquisition of this skill involves identifying "self-controlling" features of the task as a passive control strategy. Further results via local linear stability analysis showing a stable period-one solution are available (Schaal et al. 1996; Dijkstra et al. 2004).

Active control strategies are also needed to accomplish the task as intermitant contact adjustments are needed. Later examinations have revealed that participants can perform stable bouncing outside regimes of passive stability as well. Hence, participants make use of visual information to control paddle oscillations. For example, when the ballistic trajectory of the ball is perturbed by altering the energy transfer of the paddle impact, stability of the rhythmic bouncing is found faster than what a passive control would allow by altering the paddle impact in proportion to the perturbation. Therefore, both active and passive control strategies appear to work together to accomplish the task.

Many tasks resemble some involvement of passive and active control working together to achieve task goals. This appears especially the case for rhythmic or cyclic behavior that recruit hybrid transitions–discrete changes involving a contact reconfiguration with the environment. Take locomotion as an example. The human musculosketal system appears to find an extremely efficient solution for walking requiring minimal energy. Properties like dynamic stability are realized in displays of passive dynamic robots. These displays emphasize how organization of musculosketal system allow control to be "relegated down" to biomechanics rather than heavily rely on extremely complex control algorithms in the brain. In other words, this kind of design reduces the degrees of freedom needed like the motor equivalence problem emphasized.

Understanding what happens in the brain during the learning of a task that involves both passive and active control elements remains largely unknown. On the passive side, rhythmic movements have largely been explained by control circuits located in the "early brain", i.e., spinal cord and brainstem. These circuits, known as central pattern generators (CPGs), appear to be recruited in many phylogenetically early behaviors. Like the paddle juggling task itself, CPGs can easily be modeled by a nonlinear dynamical system. The generation of a sinusodal trajectory is all that is essential. Biologically plausible models like the Matsuoka half-centered neural oscillator can accomplish this (Matsuoka 2011). A sinusodal trajectory can be generated by two tonically excited neurons. These two neurons are in reciprocal inhibition of each other can easily model flexion and extension movement. How descending sensory signals from higher-order areas of the cortex appear to modulate such circuits is not known, however.

63

3.4 Objectives

The next chapter examines virtual paddle juggling as a neuroimaging investigation. Overall, the objectives of this experiment sought to perform the following:

1. Introduce the novel task of virtual paddle juggling for an online fMRI measurement.

2. Demonstrate that over the course of a single measurement participants learn to bounce a ball rhythmically and yield considerable improvement.

3. Measure and map out task-evoked patterns of group participant activity performing the motor task.

4. Using model-based investigations, determine what brain areas correlate with kinematic variables that determine the passive stability property of the task.

5. Broadly examine what are the changes of resting-state and task-evoked network patterns that evolve throughout the period of learning in a single session.

6. Determine whether a performance of the task (measured by error) is associated with measures of rsfMRI and tfMRI similarity.

Chapter 4

Dynamics of complex rhythmic skill learning with task-evoked and task-free fMRI: a virtual paddle juggling experiment

4.1 Introduction

The previous chapter sought to illuminate where previous paradigms fall short in tackling problems posed by motor skill learning. To reiterate, most neuroimaging accounts of motor control and learning emphasize the presence of a top-down, centrally organized, internally represented controller (Wolpert 1997). This account, while enormously productive, lacks emphasis on how neural representations are shaped by dynamic properties of the task and its environment (Lungarella & Sporns 2006). That is, describing how constraints imposed by them may shape behavior and consequential brain states (Warren 2006). To better examine how task and environment constraints influence formation of their neural substrates to accomplish complex motor behavior, less representational centered views have developed (Gibson 1979; Turvey & Fonseca 2009). These views are supported with the study of motor tasks rigorously definable in terms of their measured kinematics (Sternad 2017). That is, tasks that lend themselves to an analysis with concise mathematical models fully defining their dynamics. One such task is virtual paddle juggling where a ball is rhythmically bounced on a paddle to a stable, defined height goal.
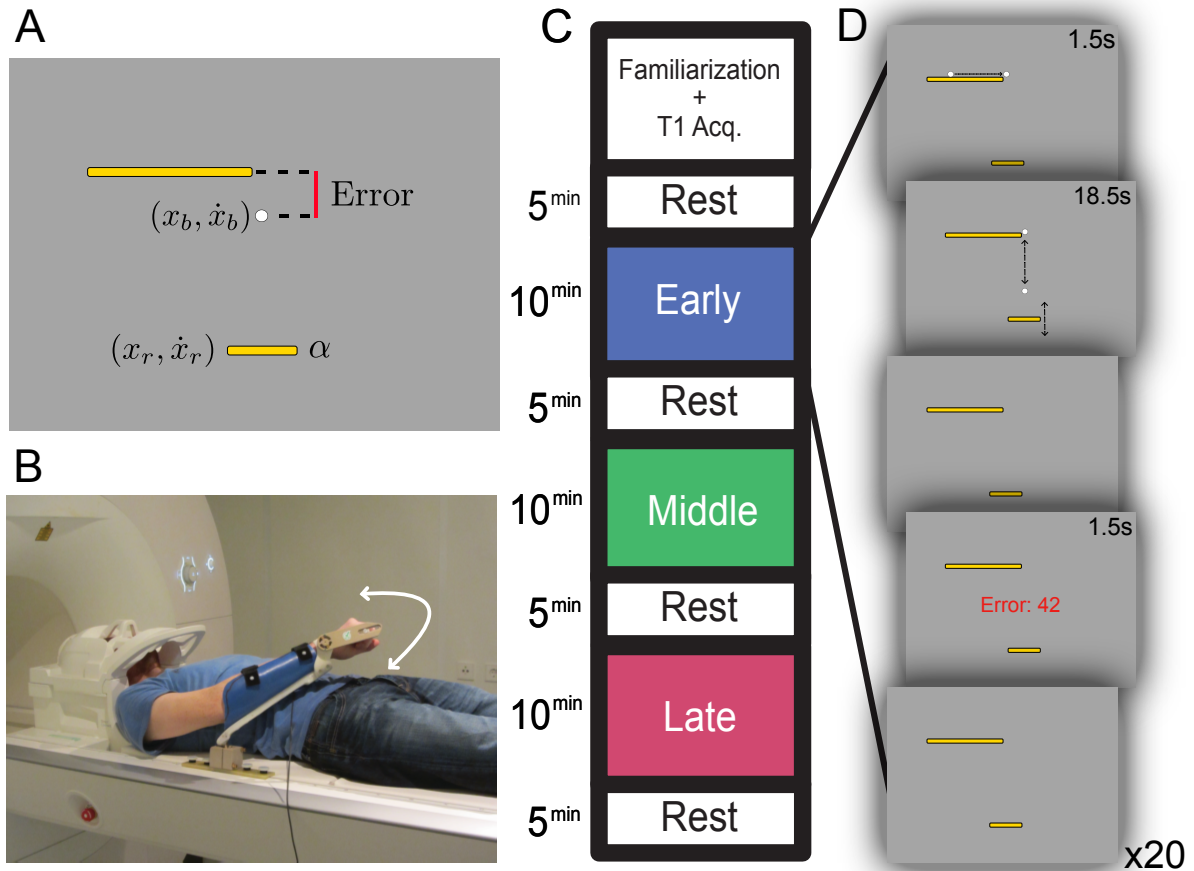
Figure 4.1: Overview of experimental setup, paradigm, stimulus, and bouncing trial dynamics - **A** Stimulus - An annotated stimulus display of the ball bouncing task and corresponding state space variables. Paddle position, paddle velocity, ball position, ball velocity are denoted by $(x_r, \dot{x}_r, x_b, \dot{x}_b)$, respectively. Errors were determined by the absolute deviation from which the ball's apex height occurred and the target height. **B** Arm-cast and environment - subjects could comfortably interact within the simple virtual environment via wrist flexions and extensions in 1 DOF. **C** Measurement session design - the experiment was composed of interleaved resting-state and task measurements each lasting roughly 5 and 10 minutes, respectively. 3 task runs, (early, middle, late), and 4 rest runs were completed in a single session lasting no longer than 1 hour. **D** Trial stimulus presentation - a single trial lasted 30 seconds with 1.5 second ball rolling, 18.5 seconds, bouncing and a explicit error feedback (median of apex deviations) during a 10 second rest period.

Behaviorally, it is known that humans performing the rhythmic task exploit dynamics of the ball-paddle interaction such that it can to large extent be "self-controlling," i.e., perturbations that would disrupt the movement will be rapidly dampened out (Schaal et al. 1996; Sternad et al. 2000; Dijkstra et al. 2004). These observations would suggest that the acquisition of this skill involves identifying "self-controlling" features of the task, i.e., a solution manifold , and then having more volitional or "conscious" control not interfere with the aspects that appear largely self-controlled (Milton et al. 2007). What is less known is what brain state or network reconfiguration would correspond to humans discovering these

kind of solutions to a task that resemble general phenomena across a myriad of complex motor behaviors involving continuous rhythmic control. Although juggling has been explored in robotics (Schaal & Atkeson 1993), human behavioral studies (Morice et al. 2007), and various theoretical investigations (Tufillaro & Albano 1986), the de novo learning of a continuous skill of this kind is yet to be explored within any detailed simultaneous neuroimaging experiment.

Here, we examine the short-term development of this skill in a single-session fMRI experiment with a specific experimental paradigm. Four interleaved resting-state (rsfMRI) between 3 task measurements (tfMRI) are acquired over the course of learning. First, behavioral data acquired simultaneous to task performance allowed a model-based investigation to dynamic state variables that sought to explain "self-controlling" aspects of the task. We hypothesized that modeling this information would reveal a correspondence to regions of the brain that modulate in accordance to performance monitoring and error correction. Second, we sought to characterize how brain networks reconfigure to explain this skill development in both rsfMRI and tfMRI. Reconfiguration in both of these two separate measurements is expected with some considerable overlap. Further, rsfMRI scans provided a basis to relate how task-evoked activity would correspond to the brain in a state where no external, goal-directed demand was present other than keep attention. To provide a measure of that relationship, we employed a recently developed model originally utilized for "connectome-fingerprinting" studies (Langs et al. 2015; Tavor et al. 2016; Cole et al. 2016; Jones et al. 2017; Tobyne et al. 2017; Tobyne et al. 2018; Osher et al. 2019; Cohen et al. 2020; Niu et al. 2020). The method is used to predict an individual's task-evoked response across different stages of learning–Early, Middle, Late. We quantify its prediction score relative to the learning scores of individual subjects as measured by their degree of error. Together, this provides a means of measuring the relative dependence between individualized task performance and how much resting-state and task-state would relate to each other. Third, we present results showing how task activity changes over the course of learning mapped over the entire cortex. These result help explain in greater detail why a correspondence between rest and task evolves during learning.

4.2   Task Model, Simulation, and Stimulus

To successfully accomplish the motor learning task, subjects bounced a virtual ball rhythmically to a fixed target line. Interaction with the virtual system was performed using a custom-built MR compatible wrist manipulandum attached to their arm and constrained to

one degree of freedom, figure 4.1 B. The task involves guiding the virtual paddle controlled via wrist flexions and extensions to impact the ball with the correct amount of energy such that the ball would not overshoot or undershoot the target line. The learning of the task involves careful wrist movement in a single direction such that the brief moments of ball-paddle impact can be controlled precisely for the duration of the trial (18.5 seconds). Since the task design involves the rhythmic bouncing of the ball, considerable visual guidance is needed to understand the behavior of how ball position errors will influence the next impact. That is, adjusting correctly for the energy needed at next impact based on the apex of ball position. The system closely resembles many everyday tasks involving perceptually guided rhythmic motion.

The dynamics of the task system are entirely definable and has undergone considerable human behavioral study lending to this property. We briefly describe the nonlinear dynamical system model used, originally developed for a particle bouncing on a vibrating surface later used in many human studies (Guckenheimer & Holmes 1983; Tufillaro et al. 1992; Schaal et al. 1996; Sternad et al. 2000). The model is a planar surface that moves sinusoidally in the vertical direction to rhythmically impact a ball. The continuous state space variables are the vertical position of the virtual ball and its velocity, $(x_b, \dot{x}_b)$, and $(x_r, \dot{x}_r)$, the paddle position and velocity, respectively. The ballistic trajectory of the ball is completely governed between the kth and the $k^{th} + 1$ impact times by the following discretized dynamical equations,

$$
\begin{aligned}
\dot{x}_{b,k+1} &= -\sqrt{((1+\alpha)\dot{x}_{r,k} - \alpha\dot{x}_{b,k})^2 - 2(x_{r,k+1} - x_{r,k})}, \\
x_{r,k+1} &= x_{r,k} + (x_{r,k+1} - x_{r,k}), \\
t_{k+1} &= \tfrac{1}{g}(((1+\alpha)\dot{x}_{r,k} - \alpha\dot{x}_{b,k}) + \dot{x}_{b,k+1}),
\end{aligned}
\tag{4.1}
$$

where t_{k+1} defines the time between successive ball-paddle impact. The loss of energy from ball-paddle impact is modeled by the coefficient of restitution α. The gravity constant of the system is defined by g. For our simulations during the scanner measurements, we choose $\alpha = 0.8$ and $g = 1.96$ (meters/second^2). These choices allowed subjects to manage a reasonable difficulty of the task and provide a realistic feel of an actual bouncing task.

Where an open-loop smooth periodic motion of the paddle and energy loss compensation are assumed, local linear stability analysis shows where the system can achieve a stable period-one solution. Indeed, (Schaal et al. 1996) first showed that at least one asymptotically fixed-point solution exists where the condition of paddle acceleration $\ddot{x}_{r,k}$ at impact is satisfied by the following inequality,

$$-2g\frac{1+\alpha}{(1+\alpha)^2} < \ddot{x}_{r,k} < 0 \tag{4.2}$$

Hence, period-one bouncing of the open-loop system is stable if impact occurs in the phase of paddle motion where acceleration is negative. If the ball dynamics are perturbed in this state space, the open-loop system will self-correct to a constant bounce height and impact acceleration. Therefore, for the system that was simulated, stable solutions were predicted if the mean paddle acceleration at impact $\ddot{x}_{r,k} \in [-2.178, 0]$.
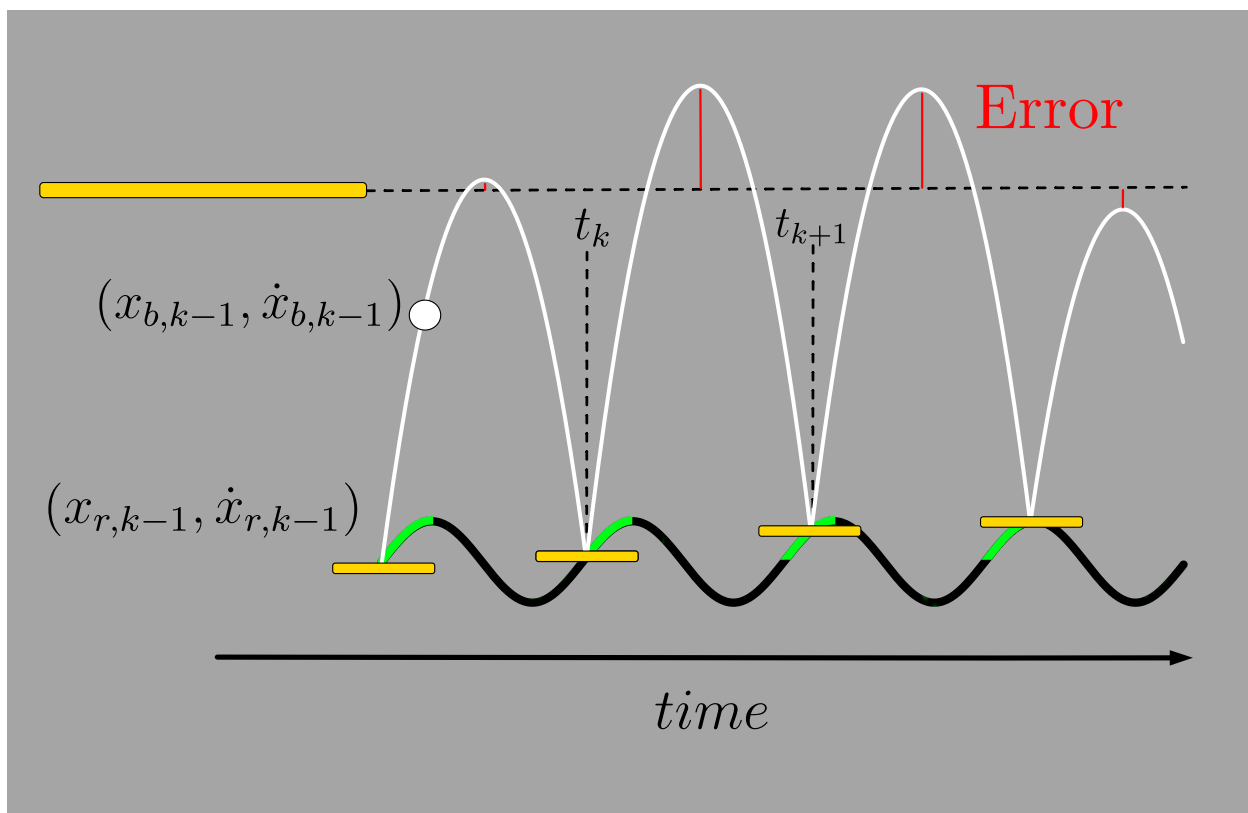
Figure 4.2: Bouncing Simulation - an example trajectory of the paddle and ball interacting with their associated state variables denoted. The manipulandum is moved in the up and down direction by wrist flexions and extensions. The green colored region of the paddle trajectory marks the passively stable of an open-loop system, i.e., where $\ddot{x}_r < 0$.

4.3 Methods and Materials

28 healthy young adults, task naive, participated (13 Female, 15 Male, mean age 27.4 ± 3.5 SD). They all self-reported to be right-hand dominant. All subjects gave written informed consent and received 12 Euro/hour as compensation. The study was approved by a local ethics committee. During scanning, subject performance was monitored online ensuring

consistency with the experimental design, i.e., no extraneous movements during resting periods. Additional inflatable head cushions (Pearl Technology, Schlieren, Switzerland) were used beyond normal foam ones to attenuate challenges of head motion during a motor fMRI experiment. An additional inflatable forehead cushion within the head coil was used to stabilize the head in the y-plane. Tape was also used over the headcoil to provide proprioceptive feedback to each subject participant in case their head position deviated (Krause et al. 2019). 3 subjects were excluded from further data analysis. Due to initial errors in data acquisition yielding incorrect volume timing information, the first 2 subjects (S1, S2) were excluded. An additional subject (S9) was excluded on the basis of excessive head motion. Therefore, 25 subject measurements were analyzed for the present study's fMRI results.

4.3.0.1 Experimental Setup and Design Overview

The overall design of the experiment used interleaved resting and task measurements. In total 4 resting-state and 3 task runs were acquired over the course of the measurement session, figure 4.1 C. A resting measurement preceded and followed every task measurement. Resting-state scans comprised of having the subject keep attention, eyes-open, and fixate on a black crosshair centered on the screen for 5 minutes, an amount of time for sufficient measurement stability considered by (Van Dijk et al. 2009). Each task scan comprised of 20 bouncing trials, each lasting 30 seconds with 18.5 seconds of motor engagement needed for the task. Feedback information was displayed after the disappearance of the ball and a randomly sampled time interval from a truncated exponential distribution, regardless of position, marking the end of the trial. No control condition was defined, however, data at 3 different times were collected for within-group comparisons. This is statistically more powerful and does not require the large number of subjects handling issues of inter-subject variance (Poldrack 2000; Thomas et al. 2009).

Participants manipulated the arm-cast joint attachment while laying in a supine position in the scanner. A simplistic 2D virtual environment was rendered on the visual display to include (1) a white ball, (2) a yellow target line, and (3) a yellow paddle, figure 4.1. Those three objects were placed upon a light-grey background. A semitransparent screen visually covering 20.67 ° y-degrees was placed at the back of the scanner bore and viewed through a mirror attached to the head coil 85cm away. The stimuli were presented onto the screen with a 1920 x 1080 pixel resolution, 60-Hz refresh rate, with a linearized projector. Participants could control the position of the virtual paddle by the moving their wrist. Stimulus presentation was implemented with PsychoPy (Peirce 2007).

Prior to scanning, all subjects underwent a familiarization session outside the scanner's bore that made clear the task goal, setup, and experiment-protocol. Subjects used their right dominant hand to perform wrist flexion and extension movements they would remain comfortable with throughout the entire experiment duration. They were instructed to not use their full range of motion to avoid haptic sensation due to enhanced contact with the handle. Additionally, subjects were instructed to bring their hand to a complete rest after the completion of each trial and not perform any extra movements in anticipation of the next trial. Compliance with this task demand was monitored. After the first resting-state scan where subjects were instructed to focus their view onto a black crosshair at the center of the screen, a 3 trial practice run was initiated to ensure the task could be achieved satisfactorily in the scanner bore. After completion of this short, familiarization phase, the first task-run was started and then followed by a short resting-state measurement instruction to have their eyes-open and stare at the centered black crosshair. Subjects were asked after completion of the measurement session what they thought about during the measurement session to interrogate whether they had any rumination of the task or any related ideas following the task-runs. All subjects reported unrelated topics, and hence the decision to not to include any additional tasks to limit such possible rumination effect in resting-state acquisitions was justified.

4.3.0.2   Behavioral Data Acquisition and Analysis

An MR-compatible arm-cast was custom-made that allowed the stabilization of the arm within a strap-tightening, adjustable plastic case. Arm position was further stabilized by additional cushions and supports. The configuration of the arm-cast was further adjusted in angle and translation position along the scanner bed. This allowed a maximally free movement within the limited size of the scanner bore relative to the participants body. It further ensures that the participant was as comfortable as possible and had a range of motion available to accomplish the task.

Data from joint angle kinematics were measured with an MR-compatible fiber-optic incremental rotary encoder (Micronor, Regensdorf, Switzerland). The encoder was attached to an arm-cast and sampled at 1000Hz, furnished with custom device drivers to its USB interface. An online linear Kalman filter estimated paddle-state position and velocity to resolve quantization errors due the high sampling frequency and finite resolution of the encoder. Data collection, filtering, the simulation of the virtual environment, and stimulus presentation were implemented to run from a single computer. To obtain an impact acceleration estimate from data acquired by the encoder that facilitated further analysis, the position signal was

differentiated and subsequently filtered by a 5th order Savitzky—Golay filter with window length of approximately 11ms. The signal was then sub-sampled to 100Hz and a 5th order smoothing spline was fit. Following this filtering, the signal was differentiated to obtain an acceleration estimate.

4.3.0.3 Image Data Acquisition and Preprocessing

All MRI measurements were performed using a 3T Siemens Prisma Trio scanner with its vendor 64-channel coil (Siemens, Erlangen, Germany) located at the Max Planck Institute MR Centre, Tübingen, Germany. A high-resolution T1-weighted magnetization-prepared rapid gradient echo anatomical scan (MPRAGE, FOV 256 x 256, 1-mm isotropic voxels) was collected at the start of the each session before any functional measurement commenced. The functional acquisition made use of a Simultaneous multi-slice T2*-weighted gradient-echo echoplanar imaging (SMS-EPI) pulse sequence optimized for temporal acceleration and sensitivity (Demetriou et al. 2018). 7 imaging runs (4 rest, 3 task) were collected, all with a multiband acceleration factor = 3, TR = 888 ms, TE = 30 ms, FA = 52 degrees, 64 x 64 acquisition matrix, 42 slices, thickness 3.0 mm with a 10% slice spacing, and an in-plane resolution 3.0 x 3.0 mm using CMRR-MB sequences (Xu et al. 2013). To aid in correcting for EPI distortions, a distortion map was estimated by an additional EPI scan with a reverse phase-encoding direction at the end of the session. To ensure T1 stabilization of functional images, the initial 5 TRs were discarded from the run. Single-band reference images acquired were used to aid in coregistration steps during preprocessing. No image artifacts were detected that could be attributed to the introduction of the arm-cast or its movement (Measured B0 Inhomogeneities, unreported in previous pilot).

Image preprocessing was performed with fMRI-Prep version 1.4.1 (Esteban et al. 2019) and is detailed in the following two sections.

4.3.0.4 Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al. 2010), distributed with ANTs 2.2.0 (Avants et al. 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (Zhang et al. 2001). Brain surfaces were reconstructed

using `recon-all` (Dale et al. 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* (Fonov et al. 2009), RRID: SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym.

4.3.0.5 Functional Data Pre-processing

For each of the 7 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. A deformation field to correct for susceptibility distortions was estimated based on two echo-planar imaging (EPI) references with opposing phase-encoding directions, using `3dQwarp` (Cox & Hyde 1997) (AFNI 20190210). Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve & Fischl 2009). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (Jenkinson et al. 2002). BOLD runs were slice-time corrected using `3dTshift` from AFNI 20190210 (Cox & Hyde 1997). The BOLD time-series, were resampled to surfaces on the following spaces: *fsaverage5*. *Grayordinates* files (Matthew F. Glasser et al. 2013), which combine surface-sampled data and volume-sampled data, were also generated. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled into standard space, generating a \emph{preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run,

73

both using their implementations in *Nipype* (Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (Behzadi et al. 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e.~head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.5.2 (Pedregosa et al., 2014), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation.

Smoothing was applied to all measurements via edge-constrained volume smoothing with a FWHM= 6mm Gaussian kernel (Smith & Brady 1997). The first 5 repetitions of each acquired run were discarded to ensure T1 stabilization. All subsequent rsfMRI and tfMRI analyses made use of stringent control of motion and physiological confound information. Subject-specific framewise displacement and 6 rigid-body motion parameters (3 translations

and 3 rotation), along with the first 6 PCA components of anatomical CompCor (Behzadi et al. 2007) were included as GLM regressors or separately removed orthogonal to applying a conservative low-pass filter (200s) (Lindquist et al. 2019) prior to any ECM, seed-based, or rest-task mapping analysis procedures.

Due to the complex motor and attention demands of the task, we carefully examined summary statistics with the potential to identify outlier volumes. These included standardized first derivative of the variance of the root mean squared head position change over voxels (DVARS) (Power et al. 2012) and framewise displacement (FD) of the task measurements. In all univariate GLM analysis, we censored any volumes that systematically proved to hold an estimated DVARS value above a statistical threshold according to the methodology of (Afyouni & Nichols 2018).

4.3.0.6   Univariate Model Estimation

For the volume-based analysis, a massive univariate modeling approach relied on smoothed data (FWHM= 6mm) in MNI-152 normalized space. Activity estimates were computed for these filtered functional time series from each run using a general linear model (GLM) fit with ordinary least squares. The following deconvolution model was used to estimate brain areas that responded to the task demands of bouncing trials: trials were modeled as boxcar regressors that encompassed the onset and duration of the bouncing trial and was one of two primary regressors of interest for contrast estimates reporting changes correlated with learning. The other primary regressor reported in model-based results was acceleration, a parametric regressor of estimated paddle acceleration at impact, a variable that described stability properties of the system, equation 4.2. Additionally, a model of ball error was examined. Impact acceleration was extracted as the moment the paddle made contact with the ball as a delta peak, while error was estimated as the median trial error. Impact acceleration and error regressors were orthogonalized with respect to the main bouncing trial regressor. To compensate for variability in the hemodynamic response function (HRF) delay across regions, temporal derivative terms derived from the predictor of interest were added to the GLM and treated as confounds of no interest. Both time series data and the GLM design matrices were temporally filtered with a Gaussian-weighted linear high-pass filter with a soft cutoff of 128s. Finally, to reduce biased ordinary least squares estimation of the GLM, a local noise model was fit to estimate the level of autocorrelation correction needed to be applied before General Linear Model (GLM) fitting (Woolrich et al. 2001; Bollmann et al. 2018). A fixed-effects model across 3 task runs was estimated. Group-level results were estimated using a permutation-based method (LISA) shown to enhance detection power over

other similarly used permutation-based methods (Lohmann et al. 2018).

4.3.0.7 Functional Correlation-based Analysis

An examination of learning related changes over the rsfMRI measurements performed involved first looking at whether any changes could be detected without defining any specific regions of interest or seeds used in conventional approaches. Therefore, Eigenvector Centrality Mapping (ECM) was used to provide a whole-brain, voxel-wise map for individual subjects (Lohmann et al. 2010). Briefly, ECM maps a value to each brain voxel. A voxel will be assigned a large value if it is strongly correlated with many other voxels that are themselves central within the network. Following ECM map computation for each run, differential changes between Late and Early runs were calculated and Z-scored.

Additionally, learning related effects in rsfMRI were tested using functional correlation-based maps via the canonical approach of selected ROIs or seeds. These ROIs and seeds were selected from the statistical parametric map of group task activation following (Abraham, Dohmatob, et al. 2014). The selected contrast map used modeled the general task condition of ball bouncing trials, see figure 4.6. ROIs were generated from contiguous clusters of voxels selected by (1) exceeding a Z-score threshold that survived aboved the 99 percentile on the data and (2) had sizes that were no smaller than $500mm^3$. Regions were parsed into contiguous clusters and 15 survived this procedure. The maximum voxel belonging in each of these clusters was selected to center a 6mm radius sphere. They were used to define a time series via mean averaging contiguous voxels belonging in a 6mm radius sphere. Seed time series of each of these 15 spheres were computed for each run.

tfMRI DMN strength was calculated based on a pre-defined atlas map independent of our data. Clusters belonging to the DMN were selected from a group-level atlas of brain networks extracted from rsfMRI via Multi-Subject Dictionary Learning (MSDL) (Varoquaux et al. 2011). Nodes within the MSDL atlas defining spatial clusters of the DMN were thresholded at 50th percentile. DMN maps were calculated for individual subjects and runs via a seed time-series computed as the mean average of voxels belonging in the isolated PCC node of the group atlas. Following this calculation, correlation seed-maps were Fisher-Z transformed. DMN strength was calculated for individual subjects and runs as the mean average correlation values found in DMN node clusters shown in figure 4.4.

All rsfMRI and tfMRI differential effects in ECM or seed maps were calculated across Late-Early runs and submitted to a non-parametric and threshold-free voxelwise one-sample t-test (LISA) (Lohmann et al. 2018).
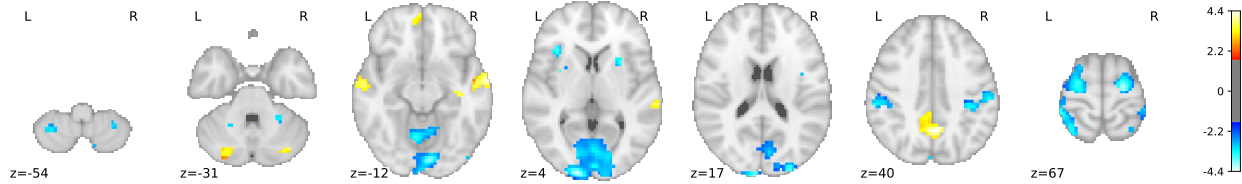
Figure 4.3: GLM map of differences between Late and Early runs of main task regressor. Primary motor regions and EVC decrease in activity, while activity within the medial frontal cortex, posterior parietal cortex, superior temporal gyrus, and cerebellar regions increase. These increases are within task-negative regions. Results are FDR corrected $p < 0.05$, 5000 permutations, at a color scale where $Z > 1.7$ and $Z < -1.7$.
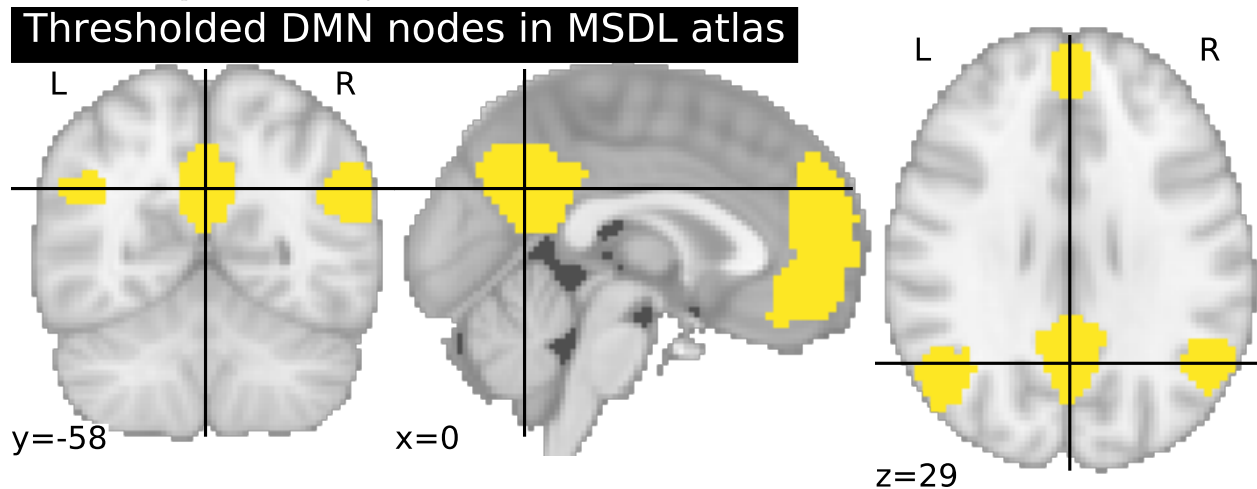

Figure 4.4: DMN-strength was calculated based on the MSDL atlas shown here (Varoquaux etal. 2011). DMN nodes were extracted and thresholded. Seed-based correlation maps were calculated based activity averaged within the PCC node centered at the crosshair.

4.3.0.8 Resting-state-Task GLM Mapping Procedure

Mapping rsfMRI to individual subject GLM maps first involves feature extraction from individual rsfMRI measurement, figure 2.1. This was accomplished on a voxel-wise level with a "semi-dense connectome" projection and is outlined here. Let $X_i \in \mathbb{R}^{v \times t}$ be subject i, demeaned, variance normalized, pre-processed and filtered resting-state measurement belonging in Naive, Early, Middle, or Late runs, where v is the whole-brain mask voxel dimension size and t the number of samples belonging in each measurement (345 TRs). All measurements across all sessions and subjects were used to first identify a group brain atlas map set of 80 components. These were calculated using a linear decomposition method initialized with Canonical ICA that enforces sparsity and non-negativity instead of independence on the spatial maps (Varoquaux et al. 2011). Let $D \in \mathbb{R}^{v \times 80}$ be the dictionary of spatial components learned from the group data. A region extraction algorithm was applied to these group-derived spatial components to retrieve 268 separate spatially contiguous clusters into matrix $A \in \mathbb{R}^{v \times 268}$ (Abraham, Fabian Pedregosa, et al. 2014). Afterwards, we applied a "dual regression" procedure to yield a separate subject-specific dictionary of spatial maps of

each subject $A_i^* \in \mathbb{R}^{v \times 268}$ (Nickerson et al. 2017). Feature extraction for individual subjects then proceeded as a calculation of a "semi-dense connectome" from subject-specific spatial maps A_i^*. Here, covariance matrix $X_i X_i^T$ is projected to form feature matrix $G_i \in \mathbb{R}^{v \times 268}$ as:

$$G_i := A_i^{*T} X_i X_i^T \tag{4.3}$$

Subject-specific features G_i were then averaged across all runs, Naive, Early, Middle, Late.

These features were then regressed onto task-GLM maps of the main ball bouncing task regressor belonging in the acceleration model and for each run session (Early, Middle, Late) independently.

Model fits were performed via leave-one-out cross validation (LOOCV) and for each voxel $j = \{1, ..., v\}$ separately. We apply a ridge regression model fit over all train subjects $n_{subjects} = 25 - 1$ as

$$\hat{w}_j := argmin_{w_j} ||\mathbf{y}_j - \mathbf{G}_j w_j||_2^2 + \lambda_j ||w_j||_2^2 \tag{4.4}$$

where $\mathbf{y}_j$ is an n-dimensional vector of task activation belonging to voxel j. The values λ_j were chosen via a generalized cross-validation procedure over the training-set data (Golub et al. 1979). Predictions scores were computed the Pearson R correlation coefficient. This was done over either whole-brain, task-negative, or task-positive regions. Task-negative and positive regions were selected based on earlier group results of the main task regression shown in figure 4.6. The fitted model is denoted by the label DL-DR-RR–Dictionary-learning, Dual-regression, Ridge-regression model.

4.3.0.9 Examining a Learning Fingerprint

In order to better localize plasticity effects, we hypothesized that differential effects over the course of learning should be reflected in *both* resting-state and task-evoked activity. That is, if the brain is indeed physically altered by some underlying mechanism of plasticity. Concretely, this means that predictions of changes in task-evoked activity should be better reflected in a model relying on *changes* in resting-state measures than in a naive resting-state measurement before any learning takes place. Hence, a comparison between a naive rsfMRI feature model versus a differential rsfMRI model would be necessary.

We make this more precise. Let $F_{j,Late-Naive}$ denote a model fit from differential semi-dense connectome rsfMRI features from Late-Naive runs $G_{j,Late-Naive}$ to predict target

$Y_{j,Late-Early}$, contrast estimates of task-evoked activity from Late and Early runs. Let $F_{j,Naive}$ denote a model fit to the same targets $Y_{j,Late-Early}$, however, with only rsfMRI features from the Naive run–$G_{j,Naive}$. Each voxel is denoted by j. Model accuracy at j was evaluated by its generalization error over cross-validation relying on *train/test* splits. Here, this is defined as

$$L(G_j^{test}; (G_j^{test}, Y_j^{test})) = Q(F_j(G_j^{train})(G_j^{test}), Y_j^{test}), \tag{4.5}$$

where F_j is the function learned from data G_j^{train} mapping to target Y_j and Q is a loss function measuring accuracy. Since this is a regression problem, we take Q to be the squared error,

$$Q(\hat{Y}_j, Y_j) = ||\hat{Y}_j - Y_j||^2,$$

where $||\cdot||$ is the Euclidean norm. A comparison of generalization error between our two models is therefore,

$$
\begin{aligned}
Q(F_{j,Late-Naive}(G_{j,Late-Naive}^{train})(G_{j,Late-Naive}^{test}), Y_{j,Late-Naive}^{test}) \\
-Q(F_{j,Naive}(X_{j,Naive}^{train})(G_{j,Naive}^{test}), Y_{j,Late-Naive}^{test})
\end{aligned}
\tag{4.6}
$$

To identify whether a model fit with $F_{j,Late-Naive}$ performs better than a model fit with features $F_{j,Naive}$, a simple paired t-test across testing folds is employed for each voxel j. However, three statistical problems make this comparison challenging in practice for neuroimaging data. First, a model's true generalization error is not known and needs to be estimated. Small sample sizes lead to very large cross-validation variance, generating unstable and biased estimates (Varoquaux et al. 2017; Varoquaux 2018). This is especially the case with commonly used cross-validation schemes in neuroimaging for inter-subject settings, e.g., leave-one-out (LOO). Second, independence assumptions of generalization errors are violated across k-fold cross-validation since resampling of the data occurs. If this non-independence is not corrected, inflated errors are generated from statistical tests that assume strict independence of samples (Nadeau & Bengio 2000). Third, each voxel is tested independently resulting in a massive multiplicity of tests. Multiple comparison correction must be applied in order to resolve the resulting inflation of statistical errors. These three factors together produce non-trivial effects leading to inflated rates of false positives errors.

The following statistical methodology was designed to resolve these challenges. Instead of leave-one-out cross-validation, a random splitting strategy with a high number of splits leaving 20% test-set is used for this comparison. Nevertheless, repeated cross-validation is no

silver bullet; although repeated random splits can enable better estimates, data scarcity is still an intrinsic and significant problem in neuroimaging where effect sizes remain extremely small. Since estimates are dependent across repeated splits, variance is underestimated (Vanwinckelen & Blockeel 2012). To try and help these underestimates when applying statistical tests, a corrected resampled t-test is used (Nadeau & Bengio 2000). Additionally, controlling for a multiplicity of tests via False Discovery Rate (FDR) enables the detection of voxels with higher sensitivity (Lohmann et al. 2018). Lastly, model $F_{j,Late-Naive}$ should not only have significantly higher generalization accuracy than $F_{j,Naive}$, but should also *explain variance* at voxel j. Therefore, the statistical procedure to identify significantly lower generalization error ($p < 0.05$) is applied as a mask upon averaged R_j^2 scores (across folds) that identify where the model being compared, i.e., $F_{j,Late-Naive}$, actually explains variance in the *test* set. This statistical control yields plausible detection of plasticity loci while controlling for false positives.

4.4   Results

4.4.1   Behavioral Learning - Task Performance

Figure 4.5 summarizes behavioral results obtained from the 3 task runs (60 trials total) performed. On average, measured subject performance demonstrates learning occurs over the single measurement session that has taken place in the scanner bore, following a typical exponential skill learning result. A one-way ANOVA with repeated measures on absolute error between training–Early, Middle, Late–reveals a significant effect of run ($F(2,54) = 24.99, p < 0.001$). Post-hoc one-sided paired t-test (5000 permutations) show significant error decreases between Early and Middle and Early and Late ($p = 0.003, p < 0.001$, respectively). This demonstrates a motor learning experiment involving interaction with a continuous dynamical system can be brought into the scanner to show short-term motor learning. Measured acceleration impacts also follow a similar shape to how error decreases. A one-way ANOVA with repeated measures on impact acceleration between training runs reveals a significant effect ($F(2,54) = 26.23, p < 0.001$). Post-hoc one-sided paired t-test (5000 permutations) show significant impact acceleration decreases between Early and Middle and Early and Late ($p = 0.001, p < 0.001$, respectively). These behavioral results show that subjects learn to perform the task within or near the dynamically stable regime shown in the shaded region of figure 4.5 B,D. Together, this suggests that finding the dynamically stable solution is an effective strategy for performing the task well.
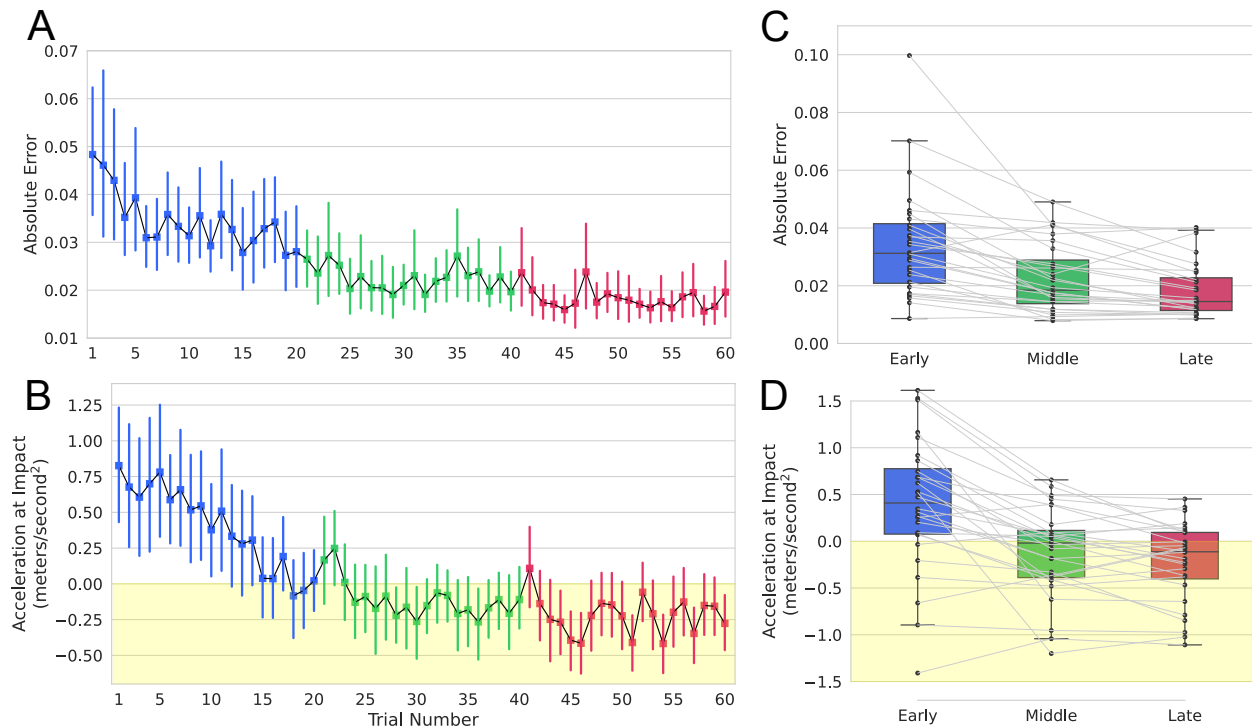
Figure 4.5: Behavioral data summary. Runs Early, Middle, Late marked by the colors blue, green, and magenta, respectively. **A, B** Trial-wise error and acceleration summary, respectively. Statistics computed are the median absolute error and median impact acceleration with 95 percent confidence intervals. Different runs across the session are denoted by 3 different colors. **C,D** Run-wise error and acceleration summary, respectively. Box and whisker plot with individual data overlaid. Yellow shaded regions in **B,D** denote regimes of passive stability.

4.4.2 Brain Imaging

4.4.2.1 Brain Activity during Task

Given the overall novelty of the task, we report task-evoked activity from measured performance of the bouncing trials. Brain activation maps were calculated via canonical massive univariate general linear model regression of the main task performance regressor. The 60 trials across Early, Middle, Late runs for individual subjects were combined for a first-level fixed-effects analysis. Figure 4.6 shows group identified brain areas with significant fMRI response elicited during the individual bouncing trials on both volumetric and surface spaces. Reported are Z-scores surviving non-parametric inference performed ($Z > 1.7$, 5000 permutations, $p < 0.05$, FDR corrected, $n = 25$) (Lohmann et al. 2018). Task related activation modeling the juggling trials revealed strong and widespread activation primarily in sensori-motor networks. These included, among other regions, the visual hMT+/V5 complex, large motor networks, mid-brain, thalamus, and lateralized brainstem activity. Additionally, the task-negative network is plotted ($Z < -1.7$, 5000 permutations, $p < 0.05$, FDR corrected,

$n = 25$). Consistent deactivations in the default-mode-network are clearly seen while the task is performed.
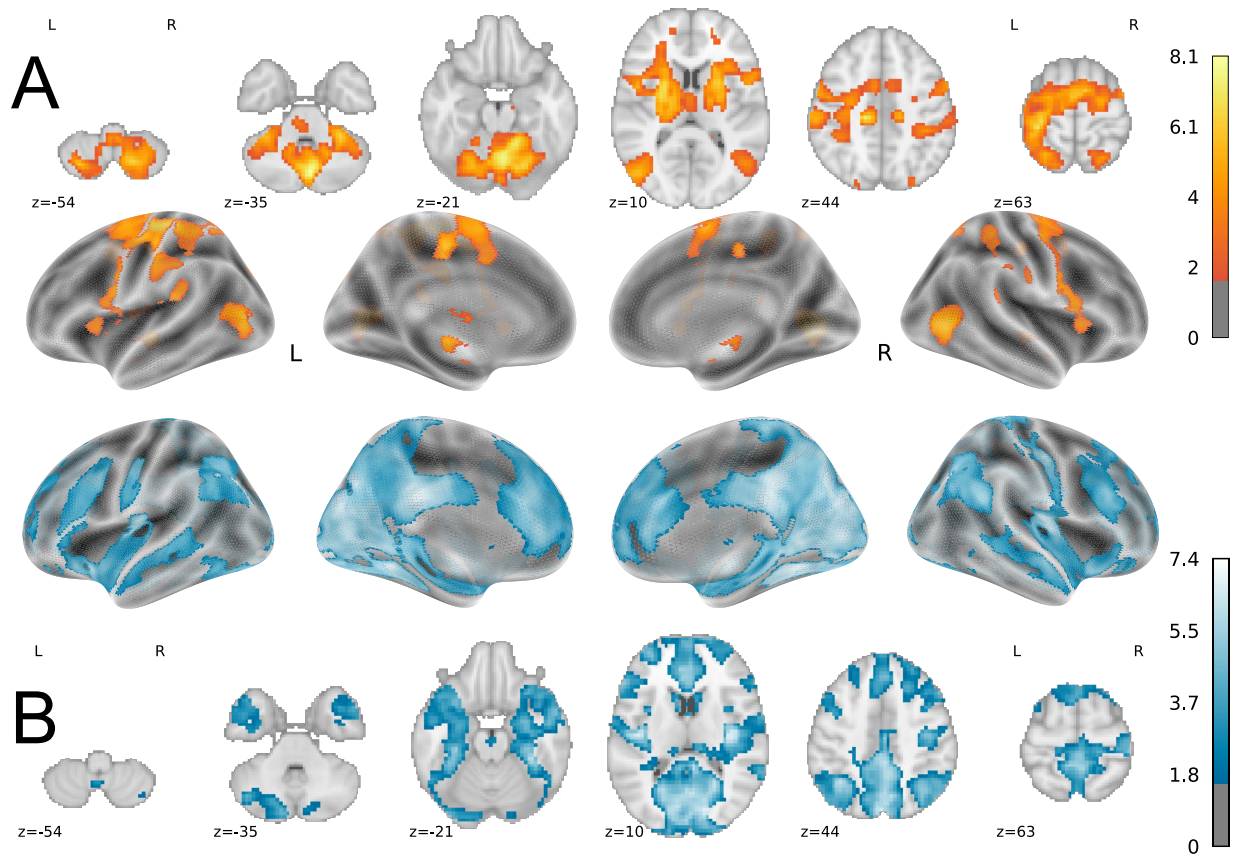

Figure 4.6: Areas recruited during main task performance, i.e., task engagement during ball bouncing trials. Results shown are volume-based mixed-effects, plotted on Freesurfer's average surface anatomy, FDR corrected $p < 0.05$, 5000 permutations, at a color scale where $Z > 1.7$ (A). Widespread activation across sensorimotor networks, cerebellar areas, and vision area are elicited. In the bottom half of the figure, the task-negative network is plotted at a color scale where $Z < -1.7$ (B).

4.4.2.2 Model-based fMRI of Task Dynamics

We hypothesized that a model-based account of both trial error and paddle impact acceleration to brain activity would reveal sensitive BOLD modulations within areas of the cortex associated with error processing and performance monitoring. To investigate this, we regressed BOLD activity of each trial onto calculated trial error and impact acceleration using a mass-univariate modelling approach, estimating the relationship between the magnitude of these measures across the 3 task runs. These analyses were performed separately for two models that either modeled acceleration or error separately.

Importantly, in both models, acceleration and error, we see significant activation modulation
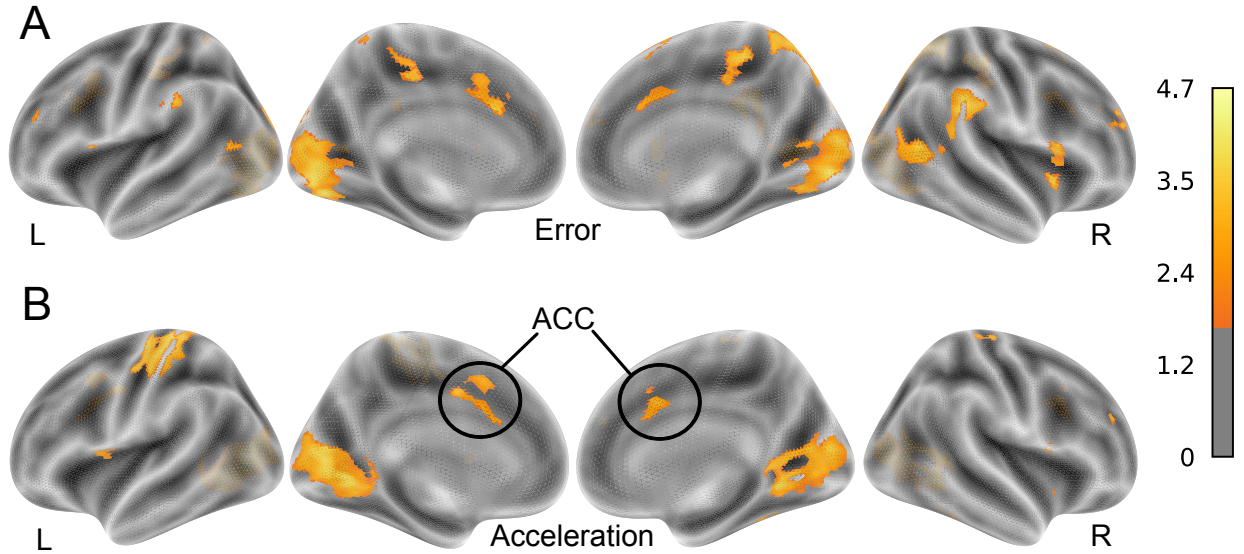
Figure 4.7: (A) Statistically significant GLM map of areas showing a parametric effect to paddle impact acceleration produced during the bouncing trials. (B) Statistically significant GLM map of areas showing a parametric effect of error according to a model akin to that shown in (A). Both models show considerable overlap in the dACC and EVC. Results shown are volume-based mixed-effects, plotted on Freesurfer's average surface anatomy, FDR corrected $p < 0.05$, 5000 permutations, at a color scale where $Z > 1.7$.

within the early visual cortex (EVC) and dorsal anterior cingulate cortex (dACC) bilaterally, figure 4.7 A, nd B. Additionally, as one intuit, the primary motor regions elicited by the task are expected to scale in response to acceleration as stronger movement is generated. Differences between these two models can largely be explained by that effect. Additionally, activity in the ventral parietal cortex (VPC) and inferior frontal gyrus (IFg) distinguish the error model compared to the acceleration one. Activity in the error model reveals the right hemisphere dominant ventral frontoparietal network. Yet, both models have similar activation profiles within the early visual cortex and dACC.

4.4.3   Relating Resting-state and Task-evoked Activity

The motivation of our experimental setup and design (shown in figure 4.1) that held interleaved rsfMRI measurements between each task run sought to examine changes across both rsfMRI and tfMRI measurements. First, we determined what kind of reorganization changes occurred due to the short-term learning within rsfMRI measurements. Next, we utilized a technique for relating rsfMRI to task-evoked activity called "connectome fingerprinting" to create a relative measure of individualized dependence between the two separate measures. The conceptualization of this is described in figure 2.1. Last, we sought to understand what kind of reorganization occurred over the course of learning within task-evoked activity. We

summarize each of these analyses results in the sections below.

4.4.3.1 No or Weak Evidence for Reorganization over rsfMRI

We investigated whether any changes across the learning session evolved within the rsfMRI measurements. Reorganization of the brain measured by rsfMRI can be supported by activity correlation changes measured before and after learning. Here, we assessed whether any observable short-term effects were present in rsfMRI patterns that was agnostic to any seed or ROI choice. Initially we took an approach that was (1) model-free; (2) voxel-wise, whole-brain, yet computationally efficient; (3) did not require choosing a selection of any specific seed regions or regions of interest (ROIs). Therefore, we apply Eigenvector Centrality Mapping (ECM) (Lohmann et al. 2010), a method that does not require any a priori model specification, e.g., a seed voxel or region of interest (ROI) or specific model parameterization, yet captures hubs in functional networks of whole-brain interactions. The method was already shown to generate support of rsfMRI changes across learning (Taubert et al. 2011). Yet, in our data, no centrality clusters compellingly appeared as differential effects between runs, neither after multiple comparison correction nor with liberal thresholding. These results suggest that no observable effects related to rsfMRI reorganization due to short-term learning were present. In an exploratory analysis, we examine functional correlation-based maps using the canonical approach requiring ROIs or seed specification. This seed-based correlation mapping sought to examine the task specificity of changes in rsfMRI due to activity generated during the motor task. Therefore, seed areas were selected from group results showing brain regions active during the performance of the task as shown in figure 4.6. Among 15 seeds isolated, two appear as statistically significant effects (uncorrected) where increased correlations with the selected seed between Late and Naive stages of resting-state measurement. However, after correcting for a multiplicity of tests (Bonferroni correction), they do not survive appropriate thresholding. Moreover, no consistent effect appears between earlier rsfMRI runs, e.g., Middle-Naive or Early-Naive. Those two significant maps are therefore unreported due to their resulting high false positive probability. Together, these two analyses suggest lack of evidence to conclude any rsfMRI reorganization associated with learning occurs.

4.4.3.2 Resting-state to Task-evoked Activity Prediction and Learning Fingerprinting
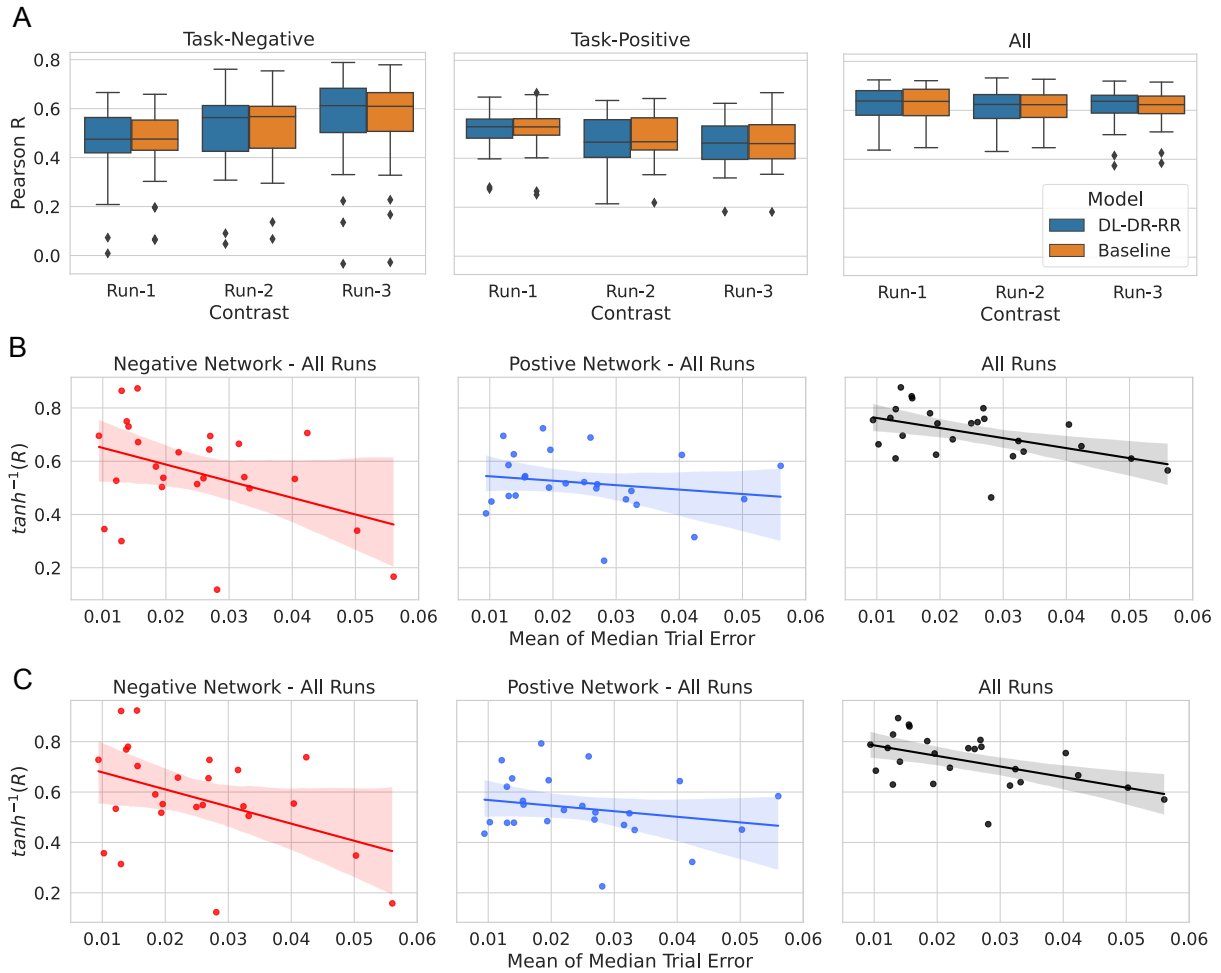
Figure 4.8: (A) Pearson R scores for individual task-evoked activity predictions summarized across Early, Middle, and Late bouncing trials between the fitted model DL-DR-RR and the baseline model. Significant increases were observed for denote either task-negative, task-positive, or whole-cortex predictions, respectively. This was in both models and they are nearly indisguishable from each other. Task-negative regions shows increasing prediction score performance while task-positive regions shows decreasing. (B) Model DL-DR-RR: individual prediction scores appear higher for participants that perform better at the task for model. (C) Model Baseline: scores

We sought in a principled way (without the selection of any seed-regions) quantify the dependence between rsfMRI and patterns of evoked activity during the bouncing task and investigate whether this had any relationship to individual task performance. To do so, we employed a regression technique conceptualized as the graphical model described in detail available in Chapter 2's figure 4.2. Briefly, this works by first extracting rsfMRI features as a voxel-wise or vertex-wise map into a "semi-dense connectome" from preprocessed rsfMRI data. rsfMRI features are then regressed onto task-evoked activity maps of the elicited activation of the virtual paddle juggling task (see method for details). Given the benchmark analysis of methods that tackle this problem, we selected one of the best known performing model in the motor domain from results in Chapter 2. However, we apply the method on a volumetric basis instead of relying on surface analysis. Because our earlier results from the previous section suggested that significant reorganization in rsfMRI was not strongly present, all rsfMRI measurements were concatenated together to enable better estimation during a critical feature extraction step. These rsfMRI features were then finally regressed onto individual GLM Z-maps across each stage of learning (Early, Middle, Late) using a single-vertex ridge regression model fit via leave-one-out cross-validation (LOOCV).

Prediction scores were quantified as Pearson correlation coefficients (Pearson R Score) between predicted GLM maps and actual individual GLM maps. This measure sought to capture whether the overall shape of the response was similar to the predicted map and summarize accuracy effectively for single subjects. We separated prediction scores evaluation scores based on either a whole-brain evaluation and divided this between a task-negative or task positive network. Task-negative areas and task-positive areas were selected from Group results shown in figure 4.6 where activation is thresholded at either $Z > 1.7$ or $Z < -1.7$ for positive and negative regions, respectively.

Figure 4.8 A shows individual prediction score results across negative, positive and whole-brain areas. Two trends appeared. First, task-negative regions relate closer to task-evoked activity across later stages of learning. A one-way repeated measures ANOVA of task-negative Pearson R Scores at Early, Middle, and Late (mean $= 0.44, 0.49, 0.54$, respectively) shows a significant effect of run, $F(2, 48) = 7.26, p = 0.0018$. Post-hoc one-sided paired t-test (5000 permutations) show significant increases between Early and Middle and Early and Late ($p = 0.032$ and $p = 0.002$, respectively). Second, task-positive regions appear to dissociate from rsfMRI activity in later stages. A one-way repeated measures ANOVA of task-positive Pearson R Scores at Early, Middle, and Late (mean $= 0.51, 0.47, 0.45$, respectively) shows significant effect of run, $F(2, 48) = 10.17, p < 0.001$. Post-hoc one-sided paired t-test (5000 permutations) show significant decreases between Early and Middle and Early and Late

($p = 0.01$ and $p < 0.001$, respectively). A one-way repeated measures ANOVA of whole-cortex Pearson R Scores at Early, Middle, and Late (mean $= 0.63, 0.60, 0.61$, respectively) does not show a significant effect of run, $F(2, 48) = 1.79, p = 0.18$.

Additionally, we investigate whether the degree of dependence between rsfMRI and tfMRI activity within individuals would relate to their behavioral performance in the task. Pearson R correlation coefficients were calculated between mean averaged prediction scores across the 3 learning session runs from the predictions we performed to mean averaged median trial error across all trials. Results are plotted in figure 4.8 B showing a significant relationship between negative-region and whole-brain Pearson R scores and averaged task performance ($p = 0.02$ and $p = 0.01$, respectively, 5000 permutations). No such relationship akin to this analysis was found for impact acceleration.

Importantly, and as results from Chapter 2 make clear (figure 2.15), we also compare the fitted model to a naive baseline model to determine the novelty of the fitting. In this case, the baseline model is taken to be the mean target samples across training subjects used in the LOOCV procedure. Previous results suggested that for some contrast targets no predictive model provides any added benefit over a simple correlation to mean activation. Figure 4.8 A,C provides similar results, showing nearly indisguishable results from the fitted model. Therefore, the fitted model appears to have no utility beyond the naive baseline model.
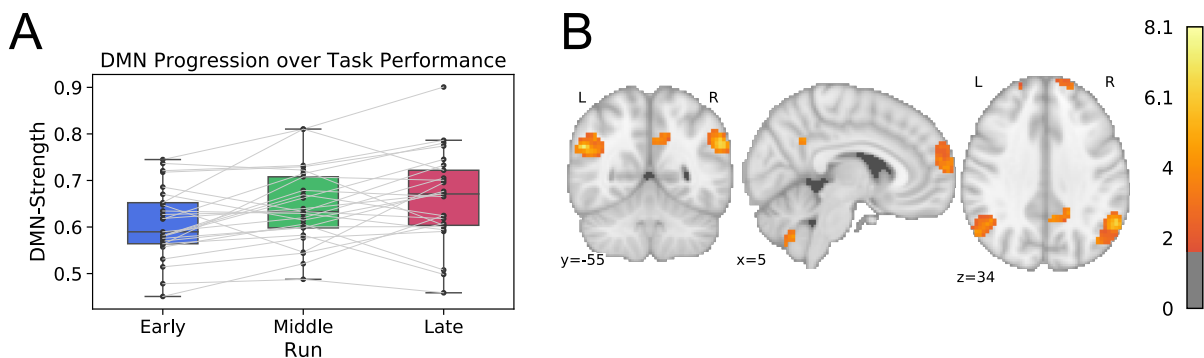

Figure 4.9: (A) Measured task-evoked DMN-strength as box-and-whisker plot over runs. Individual subject data are overlaid. Monotonic increases in DMN-strength occurs over learning runs, the highest appearing in Late. (B) Late-Early differential whole-brain changes in PCC seed-maps. The DMN prominently appears with activity in the angular gyri appearing most strongly. Additionally, a region within the cerebellum appears to be correlated with DMN activity. Results are FDR corrected $p < 0.05$, 5000 permutations, at a color scale where $Z > 1.7$.

Lastly, although lack of rsfMRI reorganization from the section earlier undermines mapping a learning fingerprint, it does not outright dismiss its possibility. Conventional methods require the specification of seed-areas, which might lead to overlooking sites where reorganization

occurs. Further, the whole-brain evaluation results might obscure small loci where rsfMRI to tfMRI prediction is actually possible. We therefore examine spatially resolved predictability with R^2 score to investigate potential sites of learning fingerprints. That is, we investigate reorganization in rsfMRI to predict reorganization detected in tfMRI using the procedure outlined in section 4.3.0.9. However, following the statistical procedure to control false positives yielded no evidence to support the presence of an identifiable learning fingerprint.

4.4.3.3 Reorganization of Task fMRI

Given the results shown in figure 4.8 within the rest-task comparison, it was evident a significant degree of reorganization is taking place in task-activity across task-performance phases. There, task-negative regions increasing relate to task-evoked activity across later stages of learning shown in both baseline and fitted models. To map these effects in detail, we first employ a GLM-based comparison between Late and Early stages of task performance. Clusters surviving thresholding ($Z > 1.7$) and multiple comparison correction (FDR, $p < 0.05$, 5000 permutations) show significant modulation within the posterior parietal cingulate (PCC), an area known to be central in the default mode network (DMN) of the brain (Raichle 2015) (See figure 4.3). GLM results suggest stronger activity in the precuneus, frontal cortex, and cerebellum in the Late stage of learning while reduced activity appears across primary visual, motor, supplementary motor regions, figure 4.3. We therefore hypothesized that an increased prominence of DMN would accompany later stages of the task performance. To measure this, we calculate functional correlations with a cluster selected a priori and isolated within the PCC from nodes making up the DMN available from a probabilistic atlas (Varoquaux et al. 2011). These nodes are shown in figure 4.4.

Measuring the strength of the DMN across runs reveals increasing utilization as measured by co-deactivations. A one-way repeated measures ANOVA was used to compare the effect of run on tfMRI DMN-strength measured at Early, Middle, and Late (mean $= 0.61, 0.64, 0.66$, respectively). Correspondingly, to examine whether habituation to the scanner environment could explain this effect, the same test was performed on rsfMRI DMN-strength measured at Naive, Early, Middle, and Late (mean $= 0.58, 0.57, 0.56, 0.57$, respectively). tfMRI results show a significant effect, $F(2, 52) = 7.57, p = 0.0014$, while rsfMRI results do not, $F(3, 78) = 0.552, p = 0.649$. Post-hoc one-sided paired t-test (5000 permutations) comparisons of tfMRI DMN-Strength show statistically significant increases from Early to Middle and Early to Late learning ($p = 0.008$ and $p = 0.003$, respectively). These results indicate stronger DMN recruitment consistent across task runs. This phenmenon appears as a result of weaker deactivation in the PCC. Together they accompany increases in task performance.

4.5   Discussion

The present study was designed to investigate how learning performance of complex rhythmic motor behavior evolves in the human brain within a single fMRI learning session. This was accomplished by utilizing a longitudinal design measuring two (typically) separately treated paradigms: rsfMRI and tfMRI. Five main findings make up our results. These are: (1) Behavioral data suggest subjects learn to perform the task in a dynamically stable regime requiring less active control. (2) Model-based fMRI results suggest acceleration–the state variable determining the open-loop stability of the dynamical system–is correlated to activity modulations occurring in error processing regions of the brain. (3) Weak to no evidence supports reorganization over short-term learning is observable in rsfMRI measurements. (4) Measures of rest-task correspondence via connectome-fingerprinting provide no utility over a naive baseline model. Further, the attempt to detect a learning fingerprint is unsuccessful. Yet, a divergence between negative and positive task-evoked regions appear. Also, these individual predictions scores relate to behavioral performance of the task. (5) Intensifying recruitment of the DMN appears as the primary driver of task-evoked network reconfiguration. Each of these results are discussed in detail below.

Prior behavioral work demonstrates subjects learning to bounce a ball rhythmically do so in a space requiring less active control, tuning into dynamical stability properties of the specific task. More precisely, under rhythmic bouncing where the periodic trajectory of the paddle is stable, participants learn to interact with the ball while instantaneous paddle impact has negative acceleration. Should paddle movement continue in an open-loop configuration, linear stability analysis shows perturbations disrupting the movement would diminish and the ball converges to its original height. Hence, participants learning to exploit this property of the dynamic system would principally require less active control. Initially, however, it was unclear whether behavior for subjects laying supine using the custom wrist flexion/extentions setup controlling the virtual paddle in an MR environment would resemble earlier experiments. This configuration was vastly different than setups of previous experiments that involved different end-effector movements to accomplish the task and for participants to be situated upright (Schaal et al. 1996; Sternad et al. 2000; Dijkstra et al. 2004; Ankarali et al. 2014). Our experimental results show, however, that participants do indeed tend to learn the task towards dynamic stability regime supporting the initial behavioral predictions generated by previous studies (Schaal et al. 1996; Sternad et al. 2000; Dijkstra et al. 2004). Decreases in acceleration also correlate with increases in task performance, i.e., minimization of error to target ball height. Together, these results demonstrate that participants learn to considerably improve over the course of a single learning session. Our behavioral data sug-

gest that one route they do so is via finding a stable state of the dynamical system defining the task.

Subject behavior supports the notion that less active control is needed to accomplish the task. Therefore, we investigated what brain areas would modulate with changes in acceleration that define the aforementioned system property of stability. Our model-based fMRI analyses show that regions known for error correction and monitoring are significantly associated with paddle impact acceleration. Investigating an akin model under error signal explicitly–instead of acceleration–reveals much of the same regions involved in these processes being modulated, figure 4.7. Namely, the dorsal anterior cingulate cortex (dACC) and early visual cortex (EVC). The dACC in particular is known to play an active role in signaling errors while engaged in a task. Broadly, activity within this region is known to be a consistent neural indicator of cognitive control mechanisms (Ridderinkhof et al. 2004; Botvinick et al. 1999), Additionally, the ACC has also been implicated to provide a means in updating internal models of the external world (O'Reilly et al. 2013). With known anatomical connections to the spinal cord and motor cortex (Paus 2001) and known influence on altering motor response (Devinsky et al. 1995), we can speculate that, like cognitive internal models, motor control ones may similarly be altered and updated (Kawato 1999). However, examinations beyond these cognitive studies, focusing on a role in motor control and learning in humans, remain sparse (Asemi et al. 2015). Nevertheless, our model-based results suggest a general link to how impact acceleration modeled in the bouncing task specifically explains how error correction and monitoring is modulated to form an active control process needed along side a passive one (Siegler et al. 2013). The error model results show modulation of the frontal parietal network, which includes the temporoparietal junction and inferior frontal gyrus, suggesting that the active control interruption of the prior smooth paddle trajectory can be considered akin to a reorienting response involving two systems (Corbetta & Shulman 2002; Corbetta et al. 2008). That is, active control is responisble for reorienting and producing a response to error information by having the ventral network interupt the passive control present in activity within the dorsal network. These results support the role of how two different systems interact to accomplish mixed control in a regime combining passive stability and active stabilization to successfully perform bouncing throughout the trial.

Our study broadly investigated general spatial reconfiguration patterns of whole-brain data occurring over the course of learning. Our experiment focused specifically on how networks within both rsfMRI measurements and task-evoked activity change and relate to one another. First, examining evidence for reconfiguration of rsfMRI activity shows very weak to no support; initial ECM analysis that did not require any a priori model specification

yielded no significant changes. This contrasts with a considerable number of other longitudinal studies that investigate alleged plasticity related effects detectable within rsfMRI. For instance, reorganization within rsfMRI longitudinal studies have been widely reported across the literature in cognitive (Nierhaus et al. 2019), visual (Urner et al. 2013), and motor learning domains (Albert et al. 2009). In our examination, only weakly supported post-hoc seed-region analysis within two areas showed significant results showed results that became insignificant after correcting for multiple comparisons. Many longitudinal examinations of motor learning examining differential changes between pre-and-post-learning phases rely on this analysis method (Vahdat et al. 2011; Vahdat et al. 2014; McGregor et al. 2018; Bernardi et al. 2018). However, statistical concerns of the post-hoc analysis lead us to conclude that evidence for this resting-state reconfiguration remains at best very weak. Additionally, the overall sensitivity of rsfMRI results to effects like time-of-day or adaptation to the scanner environment reinforce that suggestion (Steel et al. 2019). Lack of reconfiguration detectable in rsfMRI has also been reported in previous investigations involving stroke patients despite large behavioral changes (Nijboer et al. 2017; Branscheidt et al. 2019). These results emphasize that not all behavioral improvements, despite being very large, need to involve cortical reorganization—what functional connectivity methods intend to measure. The recruitment of residual cortical descending pathways, like in the case of stroke rehabilitation (Branscheidt et al. 2019), may be more relevant in the process of generating behavioral improvements.

Beyond examining rsfMRI and tfMRI reconfigurations independent of one another, we investigated how rsfMRI and tfMRI related to each other using a machine learning approach developed earlier. This was done via "connectome-fingerprinting" (Langs et al. 2015; Tavor et al. 2016; Cole et al. 2016; Jones et al. 2017; Tobyne et al. 2017; Tobyne et al. 2018; Osher et al. 2019; Cohen et al. 2020; Niu et al. 2020) that built a mapping by LOOCV to tfMRI GLM results from features derived from individual subject's rsfMRI measurement. These prediction scores provided concise summaries of rest-task dependence that could directly relate with individual measures of behavioral performance of the task as outlined in Chapter 2. We expected that task-evoked activity would increasingly resemble resting-state activity as more automated processing of the motor task became established. Additional cognitive effort would be reflected in how task-evoked activity departs from a baseline, "effortless" state where no external task demand is present. Further, this would be reflected in individual performance of the task. We found that prediction scores across all examinations relate to overall task performance, however, they are indisguishable from the naive baseline model, a result reported earlier in Chapter 2 for some contrast targets and their respective task performance. Examining the dependence between rest and task reveals a divergence

between task-positive regions and task-negative regions present in both models. Namely, prediction scores of the rsfMRI to task-evoked model show increases over learning phases within task-negative regions while decreases in task-positive regions elicited by the task. We investigate this effect in detail.

Considerable reconfiguration of networks in task-evoked activity was evident based on rest-task mapping results. In particular, task GLM differences between late and early learning reveals that the posterior cingulate cortex (PCC) is less deactivated in late, task measurements (supplementary figure 4.3). Task-negative activity and the PCC are both especially well-associated with the DMN (Raichle 2015). More broadly, the DMN is associated with baseline level of local neuronal activity when the brain is not engaged in a task demanding attention towards the external environment. An analysis of DMN-strength as measured by functional correlations was done within tfMRI measurement. PCC seed-based correlation maps reveal increasing strength within the DMN node similar to behavioral learning across all learning stages. Co-deactivity within the DMN is known to increase in response to environments that become increasing predictable and cognitive resources are freed up (Dohmatob et al. 2020). Our discovery suggests that subjects learn to perform the task in a more automated manner requiring less active control. This would align with behavioral results showing specifically how less active control is accomplished. First reports of a co-deactivation phenomenon accompanying learning we are aware of appear in a study of a shape identification task (Sigman et al. 2005). Additional previous reports of increased DMN recruitment based on decreased co-deactivation suggest this phenomenon appears in motor sequence learning (Kincses et al. 2008; Berlot et al. 2020) and cognitive learning tasks as well (Vatansever et al. 2017; Finc et al. 2020). In our motor task, we suspect that increased DMN strength may reflect how internal models for motor control are maintained from a top-down perspective. Evidence for this is revealed in increased cerebellar-DMN interaction shown in our examination as the cerebellum has been implicated in internal model formation for motor control and learning (Kawato 1999).

4.6   Conclusion

Our study presents the first fMRI investigation of human paddle juggling accomplished by wrist flexion/extension interaction. The task, although very simple, holds properties that do not typically get treatment in neuroimaging accounts of motor control and learning that we have studied here in this context. Namely, how aspects related to task dynamics influence behavioral outcomes in learning. Our observations suggest that the acquisition of

this skill involves identifying "self-controlling'' features of the task. These may act to drive reorganization observable in task-evoked activity. For instance, the increasing recruitment of the DMN. Yet, cortical reorganization across rsfMRI runs were not present.

These results together suggest that no strong reorganization suggestive of plasticity related effects from learning may be needed to accomplish the task. Instead, learning to accomplish a more passive control strategy, like this study's behavioral data strongly suggest, may be supported by simply recruiting an already present motor circuit capable of guiding a sinusodal paddle trajectory to stablize a ball's trajectory.

Chapter 5

Conclusion and Outlook

5.1 Summary of Contributions

Two projects make up the main contributions available in this thesis and are presented in Chapters 2 and 4. These chapters together constitute progress towards correcting the two problems/biases phrased at the beginning of Chapter 1.

First, an understanding of how internally generated activity shapes brain activity performed during a task was presented as a major source of issue in neuroscience. Further, it was noted that cognitive fMRI research primarily relies on task-averaged responses over many subjects to describe general principles of brain function. That is, there exists a large variability between subjects that is also reflected in spontaneous brain activity as measured by rsfMRI. Leveraging this fact, several recent studies have therefore aimed at predicting task activation from rsfMRI using various machine learning methods within a growing literature of connectome fingerprinting (Tavor et al. 2016; Cole et al. 2016; Jones et al. 2017; Tobyne et al. 2017; Tobyne et al. 2018; Osher et al. 2019; Cohen et al. 2020; Niu et al. 2020). In reviewing these results, I found lack of an evaluation against robust baselines that reliably supports a novelty of predictions for this task. On closer examination to reported methods, I found most underperform against trivial baseline model performances based on massive group averaging when whole-cortex prediction is considered. Further, this is despite having higher intra-subject than inter-subject prediction scores.

Chapter 2 presented a modification to published methods that remedies this problem to large extent. The proposed modification is based on a single-vertex approach that replaces commonly used brain parcellations. I further provide a summary of this model evaluation

by characterizing empirical properties of where prediction for this task appears possible, explaining why some predictions largely fail for certain targets. Finally, with these empirical observations I investigate whether individual prediction scores explain individual behavioral differences in a task. Work done on the side of model evaluation provide compelling evidence of an improved method. Benchmark comparisons of different methods and feature extraction approaches are reported to provide a better understanding of these methods.

These benchmarks made clear that earlier evaluation criteria found in previous literature were being used in a misleading way. Namely, the reporting of higher intra-subject vs. inter-subject prediction scores to be evidence of a successful prediction. Instead, this property should be understood to be a trivial fact. It is widely known that higher intra-subject vs. inter-subject dependence between separate, spatially normalized whole-brain measurements allows highly accurate subject identification from both rsfMRI and tfMRI-based measurements (Finn et al. 2015; Byrge & Kennedy 2019). To illustrate this concretely, I show that an output of an arbitrary encoding model of rsfMRI compared to tfMRI activation maps could reveal higher intra-subject correlation than inter-subject scores, preserving the dependency structure defining rsfMRI and tfMRI are both acquired from the same individual brain. Yet, the prediction can also be vastly poorer than a naive, unfitted baseline model. Further, previous evaluation measures based on supra-threshold extent transforms were found to be highly problematic. These problems are highlighted in Chapter 2 when discussing model evaluation choices and remain critical when evaluating models.

Rather than simply focusing on criticisms, in an effort to improve model evaluation measures, I provide alternative solutions and means of evaluation. This makes up a great deal of work shown in Chapter 2. An emphasis on baseline model evaluation and evaluating models based on a measure of R^2 was examined. To date, the study is the first relying on the later measure and was the basis to provide some important insights that evaded previous discussion. Namely, why certain contrast target predictions were vastly better than others. Finally, with these empirical observations, a relationship between individual prediction scores and individual behavioral performance during the task could be established. While justification for improving the prediction in its own right alone was clear, the utility of such predictions to explain behavior was lacking. The study was the first to show how a rest-task mapping could be taken as a measure of dependence to behavior data, a point that evolved into bridging towards the motor skill learning study.

The methodology, noted for its machine learning approach, provides a more powerful means to relate these measurement paradigms together. This approach promised a better alterna-

tive over a classical statistical approach where a preference for low-dimensional parametric models is common. Further, individual differences are mapped, avoiding the destruction of information caused by steps in group analysis procedures where averaging across individuals is necessary. An initial focus intended to improve these methods by a considerable margin over previously published ones. The need to do so was primarily motivated to create a new method capable of mapping effects related to brain plasticity, i.e., a learning fingerprint. The physiological basis, primarily from an understanding of brain plasticity, was reviewed earlier. There, it was argued that both tfMRI and rsfMRI provided the necessary means to detect differential effects caused by learning. Speculatively then, connectome fingerprinting methods would provide the specificity needed to spatially map out these differential learning-related effects between the two measurements. Improvement to this methodology was especially needed for my own data collection; my longitudinal learning study intended to make use of it.

A second study was undertaken and makes up Chapter 4. This study initially sought not only extend the use of the new methodology developed for mapping alleged brain plasticity effects, but contribute towards resolving a significant bias within the fMRI motor learning and control community. That is, fMRI skill learning examinations having low ecological validity, i.e., they do not resemble any kind of real-world behavior. To better examine tasks with higher ecological validity, I bring a virtual paddle juggling task to the scanner using custom hardware.

Within a single measurement participants learn to bounce a ball rhythmically and show considerable improvement over a single learning session. Reinforcing some earlier behavioral studies examining the task, behavioral analysis demonstrated that subjects learn the task in a state requiring less error correction and active control. Relating this to brain data acquired simultaneously to the performance of the task, model-based fMRI analysis was employed. These results show what areas of the brain correlate with kinematics of the system that identify this behavioral state—regions known to be involved in performance monitoring and error detection. These results align well with behavioral examinations showing humans performing the rhythmic task exploit dynamics of the ball-paddle interaction such that it can to large extent be "self-controlling" (Schaal et al. 1996; Sternad et al. 2000; Dijkstra et al. 2004).

The virtual paddle juggling experiment further examines reconfiguration over the two measurement paradigms and applies the methodology from the earlier study. rsfMRI reconfiguration based on functional connectivity measures suggested that no difference between

naive and late session measurements took place. This implicated the original idea of using connectome fingerprinting methodology to map out differential effects between task and rest measurements could not be easily applied to develop a means of mapping out a learning fingerprint. Instead, connectome fingerprinting from the earlier study was used to examine the overall relative dependence between rest and task summarized by whole-brain measures. Ultimately, however, this practice remained limited for reasons made clear in key results from Chapter 2, see figure 2.15. Nevertheless, utilizing this measure, either from baseline or fitted model, revealed that individual behavioral performance of the task was associated with rest-task correspondence. These results made clear that task-evoked activity increasingly related to resting-state features in task-negative associated regions. Investigating reconfiguration in task-evoked activity revealed why: stronger recruitment of the default mode network. Increasing DMN involvement appears associated with increases in task performance. Altogether, these results suggest that minimal reconfiguration of the cortex suggestive of plasticity effects are needed to find passively stable solutions that participants discover to accomplish the task.

5.2 Outlook

Critically, despite progress on fronts of both methodology and experimental work, many outstanding challenges and open questions remain. As mentioned earlier, no rsfMRI reconfiguration was detected in the paddle juggling experiment. Lack of these effects relying on older methodology made the earlier planned exploration using new methodology even more difficult. That is, mapping differential effects between task and rest, i.e., a learning fingerprinting. However, conducting a study showing effects in both rsfMRI and tfMRI and applying connectome fingerprinting methodology should still remain a top priority. A few important caveats became clear, especially after conducting the benchmark and learning study, that can be learned from. Together, they offer an insight about how to move forward with the research challenges that underlie how rest and task activity relate.

First, let us reexamine what the challenge of learning fingerprinting, the original motivation of the work conducted, actually continues to entail. Fundamentally, a procedure of statistically robust model comparison was sought after to identify plausible loci of plasticity effects. To identify these loci, a voxel or vertexwise comparison is performed between two models: a naive model relying on rsfMRI data acquired at the start of the measurement versus a differential effects model that relies on features derived from rsfMRI data capturing learning effects, e.g., between late and naive phases. These two models are fit to then predict

differential effects present in task-evoked activity between Late and Early learning stages like in the connectome fingerprinting study. Individual GLM brain maps are selected as a prediction target because they provide simple, established statistical summaries of activity changes over the course of learning. A model comparison in prediction performance that follows is used to then isolate loci of plasticity effects. Interpreting their changes as sources of plasticity effects proved to be difficult, however. To summarize from the earlier study, these reasons are: (1) Small sample size leads to estimates with large cross-validation variance. (2) Independence assumptions of these generalization errors are violated across k-fold cross-validation and need to be controlled for appropriately. (3) Each voxel or vertex is tested independently and resulting Type I statistical errors need to also be controlled.

Although these difficulties are a consequence that neuroimaging studies are often undermined by small samples sizes (Varoquaux 2018), simply gathering more samples is not the only means to make progress. Increasing sample sizes will improve power to detect learning changes, but acquiring larger samples is not always easy, especially within learning studies that require longer, more involved longitudinal experimental designs. In fact, median sample sizes for fMRI studies in 2015 still remain below 30 subjects (Poldrack et al. 2017). Further, increasing scrutiny is placed on brain-wide associations between fMRI and behavioral data as robust studies remain difficult to perform with needed reproducibility rigor (Marek et al. 2020).

Therefore, my first focus in this thesis work was to improve the power of this methodology considerably fully knowing that my own study with 28 participants would still be challenging. I believe there is still considerable gains to be made in these methods should their development continue. A promising way forward would be to project individual FC data into common/shared response spaces via shared response modeling or hyperalignment methods (Conroy et al. 2009; Conroy et al. 2013; Guntupalli et al. 2018; Richard et al. 2020). Further, a promising means to tackle challenging small sample-sized datasets would be to leverage information learned from large data sharing initiative, e.g., HCP, and then handling a domain shift onto small datasets acquired for more specific studies. There, modern machine learning methods performing semi-supervised domain adaptation appear promising toward this goal. Improving domain adaption of connectome fingerprinting is yet to be explored.

Other approaches should be taken to establish the validity of applying new methodology as well. For instance, no simulation study was carried out so far and this, in my own view, would have been helpful. Namely, this would have aided in determining needed properties of the data to ensure the hypothesized plasticity effects remain detectable. That is, especially in

regards to estimate needed effect and sample sizes Researchers working with BOLD data must eventually come to terms with the fact that they have no access to underlying ground truth of effects. Because of this, simulation studies still appear highly non-standard and controversial for fMRI method development (Welvaert & Rosseel 2014). In my view, however, simulation studies can still offer a useful surrogate to ensure that techniques are developed with a needed validity. Simultaneously, they do not need to be overly ambitious in terms of how realistic they need to be. In other words, simulations should not try to capture the full complexity brain activity, but instead focus on one specific property that is observed in real data and of central importance to the newly proposed methodology. Only a handful of essential properties of the data need to be recreated to produce the effect of interest. In my case, recreating differential effects observed during a longitudinal learning study would be needed. Although, BOLD data remains extremely complex and challenging to simulate, validation of new statistical techniques with simulated data would provide a bolstering of support that such differential effects caused by learning can be mapped out effectively. Further, properties like statistical error rate, bias, convergence, and robustness can be studied in detail. Without understanding these methodological properties, one is left in the dark about whether a study has sufficient effect and sample sizes to detect the individual learning effects I sought to map. In sum, embarking on the design of a larger and more ambitious longitudinal learning study would benefit in construction of its design informed by such simulations.

Second, voxel/vertexwise measures of variance explained demonstrate that connectome fingerprinting is largely ineffective in primary motor regions of the brain. This was largely informed by R^2 cortical mappings that show considerable variability in prediction scores across the 47 different contrast maps as figures 2.5 and 2.7 highlighted. Primary-sensory regions show little explainable inter-subject variance whereas association cortical regions show considerably better predictability, figure 2.8. For the Motor Right-Hand contrast, the best fitted algorithms did not achieve above a naive baseline score until roughly 60 subjects were used for training the algorithm, figure 2.11. Therefore, a learning task should be chosen such that differential changes fall squarly into regions of the association cortex. Individual differential effects could then be predicted by the model more easily. These might be tasks of highly cognitive nature.

Third, it still remains unclear to what degree results from connectome fingerprinting are due to underlying neural computations that importantly underlie behavior. Ultimately, fMRI signals generate indirect measures of neuronal activity measuring haemodynamic responses, e.g., blood oxygeneation, volume, or flow. Due to this, the specificity that a resting-state to task-evoked mapping holds to describe underlying neural computation remains ambiguous.

To resolve that outstanding ambiguity, learning fingerprinting sought to create an approach to specifically map effects as they related to behavior—learning a complex motor skill. Additionally, whole-brain measures show there appears to be some utility to the statistical dependence that a general connectome fingerprinting produces, e.g., shown in figure 2.15. Continuing along this line of research that asks specific behavioral questions should still remain a top priority. A successful learning fingerprint would provide considerable evidence that groups examining "trait-like" characteristics of individual differences in brain networks between task and rest are of neuroscientific relevance and not some artifact related to acquisition or physiological origin (Kraus et al. 2020). Recent progress remarking on rsfMRI fluctuation amplitudes depending on data acquisition or physiological related factors undermine some confidence in the later and this should also be explored further (Power et al. 2015; Viessmann et al. 2019; Chen et al. 2020). Of particular importance is investigating the role of dependence of individual vascular factors due to cortical orientation to the main magnetic field, B_0. BOLD signals stem from deoxyhaemoglobin that generate susceptibility differences scaling with vessel orientation. These susceptibility differences in and around blood vessels will depend on orientation to B_0. Due to this effect, unique signatures in images will be present based on factors related to the geometry of the cortex, which may not relate to any neuroscientifically relevant computations. Unfortunately, most variability of cortical geometry between subjects exists in regions that were identified to perform outstandingly well, e.g., association cortex.

For the 2mm 3T HCP data used in the study of Chapter 2, large signal biases on BOLD amplitude due to cortical orientations were identified to create up to an 11% signal amplitude bias (Viessmann et al. 2019). This observation implies that additional intra-subject dependencies between measurements exist after normalization irrespective of any functional organization structure due to underlying neurophysiology or patterns of cortical computations. To what degree this affects connectome fingerprinting results is unclear. Therefore, characterizing vascular influences remains a pressing issue to resolve ambiguity regarding the neuronal specificity of connectome fingerprinting.

5.2.1 Priming a study for the future of MR and learning.

Brain plasticity constitutes the neurological basis allowing humans to gain so much from so little–so little energy, time, and data (sensory information and experience)–to accomplish feats of new or adapted behavior. However, as MR scientists working to uncover the functional mechanisms of the brain underlying behavior, we do not get much from having little data. While the behavioral outcomes of sensorimotor learning are often hard not to notice,

the biological processes that underlie them remain hidden and require further investigation while keeping a behavioral view at the fore. These continued investigations will hopefully explain the biological mechanisms involved, characterize the time course of plasticity effects, and provide a description of how structure and function influence each other. More experiments involving larger datasets are needed towards this goal. Non-invasive in-vivo human measurement will undoubtedly continue to remain one of our best probes for characterizing whole-brain plasticity changes necessary to explain those factors in detail. Results presented in this thesis provide, what I believe, are some critical steps and have elaborated on some views I see need to be taken further.

I conclude by answering how I believe the field should continue. In my view, exploratory investigations into learning would best be served by an approach that acquires multiple data modalities that have already proven capable of measuring learning-related effects and incorporating a longitudinal design where each subject (with sufficient data) can function as a control condition (Thomas et al. 2009). This would extend beyond the two paradigms of BOLD fMRI investigated in this thesis–resting-state and task-evoked paradigms. The clear advantage of combining different MR modalities together in one learning study, i.e., characterizing one behavioral learning phenomenon, would provide in-vivo whole-brain imaging at MR to identify differential effects under functional, structural, metabolic, and vascular factors simultaneously (Hamaide et al. 2016). For MR, this would imply utilization of tfMRI, rsfMRI, T1 and T2 weighted images, DTI, and MRS together. Each technique is known to be influenced by many different biological factors lumped together and each alone cannot isolate a single process. Therefore, interpreting only one modality is severely drawn back by the resulting ambiguity when interpretation is demanded. Further, given the complexity of any MR data regardless of modality, it should not be surprising that each and every modality is plagued by unique methodological issues or confounds. Consequently, this results in decreasing the sensitivity to detect true effects. However, an approach where many converging lines of evidence can spatially localize an effect across a longitudinal examination is especially compelling and fruitful. Multiple modalities recorded together would allow a report of how these measures converge to agree with one another, providing an especially powerful means of identifying true learning effects.

Combining multiple imaging modalities is not the only means to provide stronger inferences parsing complex biological factors underlying differential effects seen in a longitudinal study. Densely sampling individuals, i.e., where a subject is measured for sufficiently long, often multiple times repeatedly, is an extremely powerful means to isolate individual level effects. For rsfMRI, up to 30 minutes may be necessary in order to retrieve stable functional correla-

tion estimates (Gordon et al. 2017) and many repeated tfMRI trials, far more than typically used in most studies, are needed. Many studies not only utilize sample sizes critically low for group analysis, but the amount of individual subject data is also a severely undermining factor. Dense-sampling approaches with repeated sampling of individuals across longitudinal learning studies remain a promising path forward (Poldrack 2017; Newbold et al. 2020).

These two design decisions—incorporating multiple MR modalities and dense sampling—influence the validity and reliability of a learning study, which ultimately define its strength. That is, whether the image changes over learning reflect what we believe is important to measure and whether the images estimate quantities are not due to sampling error and remain consistent.

List of papers by the author

Lacosse, E.; Martius, G., Lohmann, G.; Scheffler, K.;, Himmelbach, M.: Dynamics of complex rhythmic skill learning with task-evoked and task-free fMRI: a virtual paddle juggling experiment. In preparation.

Lacosse, E.; Scheffler, K.; Lohmann, G.; Martius, G.: Jumping over Baselines with New Methods to Predict Activation Maps from Resting-state fMRI. Scientific reports, 11.1, pp. 1-15 (2021)

Bause, J.; Polimeni, J.; Stelzer, J.; In, M.-H.; Ehses, P.; Kraemer-Fernandez, P.; Aghaeifar, A.; Lacosse, E.; Pohmann, R.; Scheffler, K.: Impact of prospective motion correction, distortion correction methods and large vein bias on the spatial accuracy of cortical laminar fMRI at 9.4 Tesla. NeuroImage 208, 116434, pp. 1 - 16 (2020)

Stelzer, J.; Lacosse, E.; Bause, J.; Scheffler, K.; Lohmann, G.: Brainglance: visualizing group level MRI data at one glance. Frontiers in Neuroscience 13, 972, pp. 1 - 13 (2019)

Lohmann, G.; Stelzer, J.; Lacosse, E.; Kumar, V.; Mueller, K.; Kuehn, E.; Grodd, W.; Scheffler, K.: LISA improves statistical analysis for fMRI. Nature Communications 9, 4014, pp. 1 - 9 (2018)

Acknowledgements

Statement of Contributions

Jumping over Baselines with New Methods to Predict Activation Maps from Resting-state (Chapter 2)

Eric Lacosse:
Conceptualization and Design - original research proposal
Methods - data handling and processing, software implementation, numerical/computing experiments
Writing - original manuscript, review and editing
Gabriele Lohmann:
PhD Supervision
Writing - review and editing
Georg Martius:
PhD Supervision
Writing - review and editing
Klaus Scheffler:
PhD Supervision
Writing - review

Dynamics of complex rhythmic skill learning with task-evoked and task-free fMRI: a virtual paddle juggling experiment (Chapter 4)

Eric Lacosse: Conceptualization and Design - original research proposal
Hardware Design/Software - design of wrist manipulandum and implementation of hardware device drivers and software
Experimental Design - image acquisition protocols, piloting, and stimulus design and implementation
Data Acquistion - subject recruitment, handling, and measurement
Methods - data handling and processing, software implementation
Writing - original manuscript, review and editing
Gabriele Lohmann:
PhD Supervision
Writing - review and editing
Georg Martius:
PhD Supervision
Writing - review and editing
Klaus Scheffler:
PhD Supervision
Writing - review and editing
Marc Himmelbach:

PhD Supervision
Experimental Design - piloting
Writing - review and editing

References

Abraham, A. et al., 2014. Region segmentation for sparse decompositions: Better brain parcellations from rest fMRI. *arXiv preprint arXiv:1412.3925*.

Abraham, A. et al., 2014. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8, p.14.

Afyouni, S. & Nichols, T.E., 2018. Insight and inference for dvars. *NeuroImage*, 172, pp.291–312.

Albert, N.B., Robertson, E.M. & Miall, R.C., 2009. The resting human brain and motor learning. *Current Biology*, 19(12), pp.1023–1027.

Amad, A., Seidman, J. & Draper, S.B., 2016. Motor Learning Induces Plasticity in the Resting BrainDrumming Up a Connection. *Cerebral Cortex*.

Ankarali, M.M. et al., 2014. Haptic feedback enhances rhythmic motor control by reducing variability, not improving convergence rate. *Journal of Neurophysiology*, 111(6), pp.1286–1299.

Arieli, A. et al., 1996. Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science*, 273(5283), pp.1868–1871.

Asemi, A. et al., 2015. Dorsal anterior cingulate cortex modulates supplementary motor area in coordinated unimanual motor behavior. *Frontiers in Human Neuroscience*, 9, p.309.

Aue, T., Lavelle, L.A. & Cacioppo, J.T., 2009. Great expectations: What can fMRI research tell us about psychological phenomena? *International Journal of Psychophysiology*, 73(1), pp.10–16.

Avants, B.B. et al., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), pp.26–41. Available at: http://www.sciencedirect.com/science/article/pii/S1361841507000606.

Álvarez-Salvado, E. et al., 2014. Functional mri of long-term potentiation: Imaging network plasticity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1633), p.20130152.

Bartels, A., Blaschko, M. & Shelton, J.A., 2009. Augmenting feature-driven fMRI analyses: Semi-supervised learning and resting state activity. In *Advances in neural information processing systems*. pp. 126–134.

Beckmann, C.F. et al., 2005. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457), pp.1001–1013.

Beckmann, C.F. et al., 2009. Group comparison of resting-state fmri data using multi-subject ica and dual regression. *Neuroimage*, 47, p.S148.

Behzadi, Y. et al., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1), pp.90–101. Available at: http://www.sciencedirect.com/science/articl

e/pii/S1053811907003837.

Berkes, P. et al., 2011. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013), pp.83–87.

Berlot, E., Popp, N.J. & Diedrichsen, J., 2020. A critical re-evaluation of fMRI signatures of motor sequence learning. *Elife*, 9, p.e55241.

Bernardi, N.F. et al., 2018. Error-related persistence of motor activity in resting-state networks. *Journal of cognitive neuroscience*, 30(12), pp.1883–1901.

Bernstein, N., 1967. *The coordination and regulation of movements.*, Oxford: Pergamon Press.

Bianciardi, M. et al., 2009. Modulation of spontaneous fMRI activity in human visual cortex by behavioral state. *Neuroimage*, 45(1), pp.160–168.

Bijsterbosch, J.D. et al., 2018. The relationship between spatial configuration and functional connectivity of brain regions. *Elife*, 7, p.e32992.

Biswal, B. et al., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4), pp.537–541.

Blakemore, S.-J., Frith, C.D. & Wolpert, D.M., 2001. The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport*, 12(9), pp.1879–1884.

Bliss, T.V. & Lømo, T., 1973. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, 232(2), pp.331–356.

Bollmann, S. et al., 2018. Serial correlations in single-subject fMRI with sub-second tr. *NeuroImage*, 166, pp.152–166.

Bonzano, L., Palmaro, E. & Teodorescu, R., 2015. Functional connectivity in the resting-state motor networks influences the kinematic processes during motor sequence learning. *European Journal of Neuroscience*.

Botvinick, M. et al., 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, 402(6758), pp.179–181.

Branscheidt, M. et al., 2019. No evidence for motor recovery-related cortical reorganization after stroke using resting-state fMRI. *BioRxiv*, p.681320.

Britten, K.H. et al., 1996. A relationship between behavioral choice and the visual responses of neurons in macaque mt. *Visual neuroscience*, 13(1), pp.87–100.

Brooks, R.A., 1990. Elephants don't play chess. *Robotics and autonomous systems*, 6(1-2), pp.3–15.

Buckingham, G. et al., 2014. Observing object lifting errors modulates cortico-spinal excitability and improves object lifting performance. *Cortex*, 50, pp.115–124.

Buonomano, D.V. & Merzenich, M.M., 1998. Cortical plasticity: From synapses to maps. *Annual review of neuroscience*, 21(1), pp.149–186.

Byrge, L. & Kennedy, D.P., 2019. High-accuracy individual identification using a "thin slice" of the functional connectome. *Network Neuroscience*, 3(2), pp.363–383.

Bzdok, D. et al., 2015. Semi-supervised factored logistic regression for high-dimensional neuroimaging data. In *Advances in neural information processing systems*. pp. 3348–3356.

Bzdok, D. et al., 2016. Formal models of the network co-occurrence underlying mental operations. *PLoS computational biology*, 12(6).

Bzdok, D. & Yeo, B.T., 2017. Inference in the age of big data: Future perspectives on neuroscience. *Neuroimage*, 155, pp.549–564.

Calmels, C., 2020. Neural correlates of motor expertise: Extensive motor training and cortical changes. *Brain research*, 1739, p.146323.

Castro, L.N.G. et al., 2014. Environmental consistency determines the rate of motor adaptation. *Current Biology*, 24(10), pp.1050–1061.

Censor, N., Sagi, D. & Cohen, L.G., 2012. Common mechanisms of human perceptual and motor learning. *Nature Reviews Neuroscience*, 13(9), pp.658–664.

Chaisanguanthum, K.S., Shen, H.H. & Sabes, P.N., 2014. Motor variability arises from a slow random walk in neural state. *Journal of Neuroscience*, 34(36), pp.12071–12080.

Chemero, A., 2011. *Radical embodied cognitive science*, MIT press.

Chen, J. et al., 2020. Resting-state "physiological networks". *Neuroimage*, p.116707.

Choi, Y., Shin, E.Y. & Kim, S., 2019. Double dissociation of fMRI activity in the caudate nucleus supports de novo motor skill learning. *bioRxiv*.

Cohen, A.D. et al., 2020. Regression-based machine-learning approaches to predict task activation using resting-state fMRI. *Human Brain Mapping*, 41(3), pp.815–826.

Cole, M.W. et al., 2014. Intrinsic and task-evoked network architectures of the human brain. *Neuron*, 83(1), pp.238–251.

Cole, M.W. et al., 2016. Activity flow over resting-state networks shapes cognitive task activations. *Nature neuroscience*, 19(12), p.1718.

Conroy, B.R. et al., 2013. Inter-subject alignment of human cortical anatomy using functional connectivity. *NeuroImage*, 81, pp.400–411.

Conroy, B. et al., 2009. FMRI-based inter-subject cortical alignment using functional connectivity. In *Advances in neural information processing systems*. pp. 378–386.

Corbetta, M., Patel, G. & Shulman, G.L., 2008. The reorienting system of the human brain: From environment to theory of mind. *Neuron*, 58(3), pp.306–324.

Corbetta, M. & Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, 3(3), pp.201–215.

Cortes, C. et al., 1994. Learning curves: Asymptotic values and rate of convergence. In *Advances in neural information processing systems*. pp. 327–334.

Cox, R.W. & Hyde, J.S., 1997. Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, 10(4-5), pp.171–178.

Curto, C. et al., 2009. A simple model of cortical dynamics explains variability and state dependence of sensory responses in urethane-anesthetized auditory cortex. *Journal of neuroscience*, 29(34), pp.10600–10612.

Dadi, K. et al., 2019. Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage*, 192, pp.115–134.

Dale, A.M., Fischl, B. & Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9(2), pp.179–194. Available at: http://www.sciencedirect.com/science/article/pii/S1053811998903950.

D'Angelo, E., 2018. Physiology of the cerebellum. In *Handbook of clinical neurology.* Elsevier, pp. 85–108.

Dayan, E. & Cohen, L.G., 2011. Neuroplasticity subserving motor skill learning. *Neuron.*

De Jong, J., 1957. The effects of increasing skill on cycle time and its consequences for time standards. *Ergonomics*, 1(1), pp.51–60.

Demetriou, L. et al., 2018. A comprehensive evaluation of increasing temporal resolution with multiband-accelerated protocols and effects on statistical outcome measures in fMRI. *NeuroImage*, 176, pp.404–416.

Der, R. & Martius, G., 2015. Novel plasticity rule can explain the development of sensorimotor intelligence. *Proceedings of the National Academy of Sciences*, 112(45), pp.E6224–E6232.

Devinsky, O., Morrell, M.J. & Vogt, B.A., 1995. Contributions of anterior cingulate cortex to behaviour. *Brain*, 118(1), pp.279–306.

Dijkstra, T.M. et al., 2004. The dialogue between data and model: Passive stability and relaxation behavior in a ball bouncing task. *Nonlinear Studies*, 11(3), pp.319–344.

Dohmatob, E., Dumas, G. & Bzdok, D., 2020. Dark control: The default mode network as a reinforcement learning agent. *Human Brain Mapping.*

Douglas, R.J. & Martin, K.A., 2004. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27, pp.419–451.

Draganski, B. et al., 2004. Changes in grey matter induced by training. *Nature*, 427(6972), pp.311–312.

Draganski, B. & May, A., 2008. Training-induced structural changes in the adult human brain. *Behavioural brain research*, 192(1), pp.137–142.

Driemeyer, J. et al., 2008. Changes in gray matter induced by learning—revisited. *PloS one*, 3(7), p.e2669.

Dubois, J. & Adolphs, R., 2016. Building a science of individual differences from fMRI. *Trends in cognitive sciences*, 20(6), pp.425–443.

Elliott, M.L. et al., 2019. Poor test-retest reliability of task-fMRI: New empirical evidence and a meta-analysis. *BioRxiv*, p.681700.

Esteban, O. et al., 2019. FMRIPrep: A robust preprocessing pipeline for functional mri. *Nature methods*, 16(1), p.111.

Finc, K. et al., 2020. Dynamic reconfiguration of functional brain networks during working memory training. *Nature communications*, 11(1), pp.1–15.

Finn, E.S. et al., 2020. Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, p.116828.

Finn, E.S. et al., 2015. Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11), p.1664.

Fonov, V. et al., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47, Supplement 1, p.S102.

Ganguly, K. & Poo, M.-m., 2013. Activity-dependent neural plasticity from bench to bedside. *Neuron*, 80(3), pp.729–741.

Gao, J.S. et al., 2015. Pycortex: An interactive surface visualizer for fMRI. *Frontiers in neuroinformatics*, 9, p.23.

Georgopoulos, A.P. et al., 1982. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, 2(11), pp.1527–1537.

Gerstner, W. et al., 1996. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(6595), pp.76–78.

Gibson, J.J., 1979. *The ecological approach to visual perception*, Houghton Mifflin.

Gilpin, L.H. et al., 2018. Explaining explanations: An overview of interpretability of machine learning. In *2018 ieee 5th international conference on data science and advanced analytics (dsaa)*. IEEE, pp. 80–89.

Glasser, M.F. et al., 2016. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615), pp.171–178.

Glasser, M.F. et al., 2013. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80, pp.105–124. Available at: http://www.sciencedirect.com/science/article/pii/S105381191300 5053.

Glasser et al., 2013. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80, pp.105–124.

Golub, G.H., Heath, M. & Wahba, G., 1979. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2), pp.215–223.

Gordon, E.M. et al., 2017. Precision functional mapping of individual human brains. *Neuron*, 95(4), pp.791–807.

Grafton, S.T., Hazeltine, E. & Ivry, R., 1995a. Functional Mapping of Sequence Learning in Normal Humans. *Journal of Cognitive Neuroscience*.

Grafton, S.T., Hazeltine, E. & Ivry, R., 1995b. Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience*, 7(4), pp.497–510.

Gratton, C. et al., 2018. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, 98(2), pp.439–452.

Greicius, M.D. & Menon, V., 2004. Default-mode activity during a passive sensory task: Uncoupled from deactivation but impacting activation. *Journal of cognitive neuroscience*, 16(9), pp.1484–1492.

Greve, D.N. & Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), pp.63–72.

Guckenheimer, J. & Holmes, P., 1983. Local bifurcations. In *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer, pp. 117–165.

Guerra-Carrillo, B., Mackey, A.P. & Bunge, S.A., 2014. Resting-state fMRI: A window into human brain plasticity. *The Neuroscientist*, 20(5), pp.522–533.

Guidotti, R., Del Gratta, C. & Baldassarre, A., 2015. Visual learning induces changes in resting-state fMRI multivariate pattern of information. *Journal of Neuroscience*.

Guntupalli, J.S., Feilong, M. & Haxby, J.V., 2018. A computational model of shared fine-scale structure in the human connectome. *PLoS computational biology*, 14(4), p.e1006120.

Gusnard, D.A. & Raichle, M.E., 2001. Searching for a baseline: Functional imaging and the resting human brain. *Nature reviews neuroscience*, 2(10), pp.685–694.

Halko, N., Martinsson, P.-G. & Tropp, J.A., 2011. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2), pp.217–288.

Hamaide, J., De Groof, G. & Van der Linden, A., 2016. Neuroplasticity and mri: A perfect match. *NeuroImage*, 131, pp.13–28.

Harmelech, T. et al., 2013. The day-after effect: Long term, hebbian-like restructuring of resting-state fMRI patterns induced by a single epoch of cortical activation. *Journal of Neuroscience*, 33(22), pp.9488–9497.

Hasson, C.J. et al., 2016. Neuromotor noise is malleable by amplifying perceived errors. *PLoS computational biology*, 12(8), p.e1005044.

Hasson, U., Nusbaum, H.C. & Small, S.L., 2009. Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences*, 106(26), pp.10841–10846.

Hazeltine, E., Grafton, S.T. & Ivry, R., 1997. Attention and stimulus characteristics determine the locus of motor-sequence encoding. A pet study. *Brain: a journal of neurology*, 120(1), pp.123–140.

Hearne, L.J., Cocchi, L. & Zalesky, A., 2017. Reconfiguration of brain network architectures between resting state and complexity-dependent cognitive reasoning. *The Journal of Neuroscience*.

Hebb, D.O., 1949. *The organization of behavior: A neuropsychological theory*, J. Wiley; Chapman & Hall.

Herzfeld, D.J. et al., 2014. A memory of errors in sensorimotor learning. *Science*, 345(6202), pp.1349–1353.

Hesselmann, G. et al., 2008. Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proceedings of the National Academy of Sciences*, 105(31), pp.10984–10989.

Honey, C.J. et al., 2007. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences*, 104(24), pp.10240–10245.

Honey, C.J. et al., 2009. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6), pp.2035–2040.

Hyvarinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks*, 10(3), pp.626–634.

Ito, M. & Kano, M., 1982. Long-lasting depression of parallel fiber-purkinje cell transmission induced by conjunctive stimulation of parallel fibers and climbing fibers in the cerebellar cortex. *Neuroscience letters*, 33(3), pp.253–258.

Jenkins, I. et al., 1994. Motor sequence learning: A study with positron emission tomography. *Journal of Neuroscience*, 14(6), pp.3775–3790.

Jenkinson, M. et al., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), pp.825–841. Available at: http://www.sciencedirect.com/science/article/pii/S1053811902911328.

Jillings, S. et al., 2020. Macro-and microstructural changes in cosmonauts' brains after long-duration spaceflight. *Science Advances*, 6(36), p.eaaz9488.

Jin, X. & Costa, R.M., 2010. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature*, 466(7305), pp.457–462.

Johnson, W.B., Lindenstrauss, J. & Schechtman, G., 1986. Extensions of lipschitz maps into banach spaces. *Israel Journal of Mathematics*, 54(2), pp.129–138.

Jolliffe, I.T., 1982. A note on the use of principal components in regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(3), pp.300–303.

Jonas, E. & Kording, K.P., 2017. Could a neuroscientist understand a microprocessor? *PLoS computational biology*, 13(1), p.e1005268.

Jones, O.P. et al., 2017. Resting connectivity predicts task activation in pre-surgical populations. *NeuroImage: Clinical*, 13, pp.378–385.

Kami, A. et al., 1995. Functional mri evidence for adult motor cortex plasticity during motor skill learning. *Nature*, 377(6545), pp.155–158.

Kannurpatti, S.S. & Biswal, B.B., 2012. Prediction of task-related bold fMRI with amplitude signatures of resting-state fMRI. *Frontiers in systems neuroscience*, 6, p.7.

Karmarkar, U.R. & Dan, Y., 2006. Experience-dependent plasticity in adult visual cortex. *Neuron*, 52(4), pp.577–585.

Karni, A. et al., 1998. The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences*, 95(3), pp.861–868.

Kawato, M., 1999. Internal models for motor control and trajectory planning. *Current opinion in neurobiology*, 9(6), pp.718–727.

Keller, T.A. & Just, M.A., 2016. Structural and functional neuroplasticity in human learning of spatial routes. *NeuroImage*, 125, pp.256–266.

Kenet, T. et al., 2003. Spontaneously emerging cortical representations of visual attributes. *Nature*, 425(6961), pp.954–956.

Khosla, M. et al., 2019. Machine learning in resting-state fMRI analysis. *Magnetic Resonance Imaging*.

Kim, D. et al., 2019. Spontaneously emerging patterns in human visual cortex and their functional connectivity are linked to the patterns evoked by visual stimuli. *bioRxiv*, p.518712.

Kincses, Z.T. et al., 2008. Model-free characterization of brain functional networks for motor sequence learning using fMRI. *Neuroimage*, 39(4), pp.1950–1958.

Kisley, M.A. & Gerstein, G.L., 1999. Trial-to-trial variability and state-dependent modulation of auditory-evoked responses in cortex. *Journal of Neuroscience*, 19(23), pp.10451–10460.

Klein, A. et al., 2017. Mindboggling morphometry of human brains. *PLOS Computational Biology*, 13(2), p.e1005350. Available at: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005350.

Körding, K.P. & Wolpert, D.M., 2004. Bayesian integration in sensorimotor learning. *Nature*, 427(6971), pp.244–247.

Krakauer, J.W. et al., 2017. Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93(3), pp.480–490.

Krakauer, J.W., Ghez, C. & Ghilardi, M.F., 2005. Adaptation to visuomotor transformations: Consolidation, interference, and forgetting. *Journal of Neuroscience*, 25(2), pp.473–478.

Krakauer, J.W. et al., 2011. Motor learning. *Comprehensive Physiology*, 9(2), pp.613–663.

Kraus, B. et al., 2020. Network variants are similar between task and rest states. *bioRxiv*.

Krause, F. et al., 2019. Active head motion reduction in magnetic resonance imaging using tactile feedback. *Human brain mapping*, 40(14), pp.4026–4037.

Krienen, F.M., Yeo, B.T. & Buckner, R.L., 2014. Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1653), p.20130526.

Kullmann, D.M., 2020. Editorial. *Brain*, 143(4), pp.1045–1045. Available at: https://doi.org/10.1093/brain/awaa082.

Laird, A.R. et al., 2011. Behavioral interpretations of intrinsic connectivity networks. *Journal of cognitive neuroscience*, 23(12), pp.4022–4037.

Lalazar, H. & Vaadia, E., 2008. Neural basis of sensorimotor learning: Modifying internal models. *Current opinion in neurobiology*, 18(6), pp.573–581.

Lanczos, C., 1964. Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, 1(1), pp.76–85. Available at: http://epubs.siam.org/doi/10.1137/0701007.

Langs, G., Golland, P. & Ghosh, S.S., 2015. Predicting activation across individuals with resting-state functional connectivity based multi-atlas label fusion. In *International conference on medical image computing and computer-assisted intervention*. Springer, pp. 313–320.

Lashley, K.S., 1951. *The problem of serial order in behavior*, Bobbs-Merrill Oxford, United Kingdom.

Lewis, C.M. & Baldassarre, A., 2009. Learning sculpts the spontaneous activity of the resting human brain. *Proceedings of the National Academy of Sciences*.

Lindquist, M.A. et al., 2019. Modular preprocessing pipelines can reintroduce artifacts into fMRI data. *Human brain mapping*, 40(8), pp.2358–2376.

Llinás, R.R., 2002. *I of the vortex: From neurons to self*, MIT press.

Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. *Nature*, 453(7197), p.869.

Lohmann, G. et al., 2012. Critical comments on dynamic causal modelling. *Neuroimage*, 59(3), pp.2322–2329.

Lohmann, G. et al., 2010. Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. *PloS one*, 5(4), p.e10232.

Lohmann, G. et al., 2018. LISA improves statistical analysis for fMRI. *Nature communications*, 9(1), pp.1–9.

Lohmann, G. et al., 2016. Task-related edge density (ted)—a new method for revealing dynamic network formation in fMRI data of the human brain. *PloS one*, 11(6).

Lungarella, M. & Sporns, O., 2006. Mapping information flow in sensorimotor networks. *PLoS Comput Biol*, 2(10), p.e144.

Magistretti, P.J., 2016. Imaging brain aerobic glycolysis as a marker of synaptic plasticity. *Proceedings of the National Academy of Sciences*, 113(26), pp.7015–7016.

Maguire, E.A. et al., 2000. Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, 97(8), pp.4398–4403.

Makino, H. et al., 2016. Circuit mechanisms of sensorimotor learning. *Neuron*, 92(4), pp.705–721.

Marcus, D. et al., 2011. Informatics and data mining tools and strategies for the human connectome project. *Frontiers in neuroinformatics*, 5, p.4.

Marek, S. et al., 2020. Towards reproducible brain-wide association studies. *bioRxiv*.

Margulies, D.S. et al., 2009. Precuneus shares intrinsic functional architecture in humans and monkeys. *Proceedings of the National Academy of Sciences*, 106(47), pp.20069–20074.

Markram, H. et al., 1997. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297), pp.213–215.

Mathis, M.W., Mathis, A. & Uchida, N., 2017. Somatosensory cortex plays an essential role in forelimb motor adaptation in mice. *Neuron*, 93(6), pp.1493–1503.

Matsuoka, K., 2011. Analysis of a neural oscillator. *Biological Cybernetics*, 104(4-5), pp.297–304.

Mattar, M.G., Thompson-Schill, S.L. & Bassett, D.S., 2018. The network architecture of value learning. *Network Neuroscience*, 2(02), pp.128–149.

McGregor, H.R., Cashaback, J.G. & Gribble, P.L., 2018. Somatosensory perceptual training enhances motor learning by observing. *Journal of neurophysiology*, 120(6), pp.3017–3025.

McGregor, H.R. & Gribble, P.L., 2017. Functional connectivity between somatosensory and motor brain areas predicts individual differences in motor learning by observing. *Journal of neurophysiology*, 118(2), pp.1235–1243.

Milton, J. et al., 2007. The mind of expert motor performance is cool and focused. *Neuroimage*, 35(2), pp.804–813.

Monti, M.M., 2011. Statistical analysis of fMRI time-series: A critical review of the glm approach. *Frontiers in human neuroscience*, 5, p.28.

Morice, A.H. et al., 2007. Learning new perception–action solutions in virtual ball bouncing. *Experimental brain research*, 181(2), pp.249–265.

Mueller, S. et al., 2013. Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3), pp.586–595.

Müller, H. & Sternad, D., 2009. Motor learning: Changes in the structure of variability in a redundant task. In *Progress in motor control*. Springer, pp. 439–456.

Nadeau, C. & Bengio, Y., 2000. Inference for the generalization error. In *Advances in neural information processing systems*. pp. 307–313.

Naselaris, T. et al., 2011. Encoding and decoding in fMRI. *Neuroimage*, 56(2), pp.400–410.

Newbold, D.J. et al., 2020. Plasticity and spontaneous activity pulses in disused human brain circuits. *Neuron*, 107(3), pp.580–589.

Nickerson, L.D., 2018. Replication of resting state-task network correspondence and novel findings on brain network activation during task fMRI in the human connectome project study. *Scientific reports*, 8(1), p.17543.

Nickerson, L.D. et al., 2017. Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Frontiers in neuroscience*, 11, p.115.

Nierhaus, T. et al., 2019. Immediate brain plasticity after one hour of brain–computer interface (bci). *The Journal of physiology*.

Nijboer, T.C. et al., 2017. No changes in functional connectivity during motor recovery beyond 5 weeks after stroke; a longitudinal resting-state fMRI study. *Plos one*, 12(6), p.e0178017.

Nissen, M.J. & Bullemer, P., 1987. Attentional requirements of learning: Evidence from performance measures. *Cognitive psychology*, 19(1), pp.1–32.

Niu, C. et al., 2020. Modeling motor task activation from resting-state fMRI using machine learning in individual subjects. *Brain Imaging and Behavior*, pp.1–11.

Northoff, G., Qin, P. & Nakao, T., 2010. Rest-stimulus interaction in the brain: A review. *Trends in neurosciences*, 33(6), pp.277–284.

O'Doherty, J.P., Hampton, A. & Kim, H., 2007. Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of sciences*, 1104(1), pp.35–53.

O'Reilly, J.X. et al., 2013. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, 110(38), pp.E3660–E3669.

Osher, D.E., Brissenden, J.A. & Somers, D.C., 2019. Predicting an individual's dorsal attention network activity from functional connectivity fingerprints. *Journal of neurophysiology*, 122(1), pp.232–240.

Ostry, D.J. & Gribble, P.L., 2016. Sensory plasticity in human motor learning. *Trends in neurosciences*, 39(2), pp.114–123.

Paus, T., 2001. Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature reviews neuroscience*, 2(6), pp.417–424.

Pedregosa, F. et al., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, pp.2825–2830.

Peirce, J.W., 2007. PsychoPy—psychophysics software in python. *Journal of neuroscience methods*, 162(1-2), pp.8–13.

Penhune, V.B. & Steele, C.J., 2012. Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. *Behavioural Brain Research*.

Pervaiz, U. et al., 2020. Optimising network modelling methods for fMRI. *NeuroImage*, p.116604.

Pessoa, L., 2014. Understanding brain networks and brain organization. *Physics of life reviews*, 11(3), pp.400–435.

Peterson, E.J. & Seger, C.A., 2018. In model-based fMRI significant is less than specific. *bioRxiv*, p.429621.

Pfeifer, R., Lungarella, M. & Iida, F., 2007. Self-organization, embodiment, and biologically inspired robotics. *science*, 318(5853), pp.1088–1093.

Poldrack, R.A., 2017. Precision neuroscience: Dense sampling of individual brains. *Neuron*, 95(4), pp.727–729.

Poldrack, R.A. et al., 2017. Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature reviews neuroscience*, 18(2), p.115.

Power, J.D. et al., 2012. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage*, 59(3), pp.2142–2154.

Power, J.D. et al., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, 84(Supplement C), pp.320–341. Available at: http://www.sciencedirect.com/science/article/pii/S1053811913009117.

Power, J.D., Schlaggar, B.L. & Petersen, S.E., 2015. Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage*, 105, pp.536–551.

Project, H.C., 2019a. HCPpipelines. *GitHub repository*.

Project, H.C., 2019b. HCPpipelines. *GitHub repository*.

Raichle, M.E., 2015. The brain's default mode network. *Annual review of neuroscience*, 38, pp.433–447.

Raichle, M.E. & Mintun, M.A., 2006. Brain work and brain imaging. *Annu. Rev. Neurosci.*, 29, pp.449–476.

Reichelt, A.F. et al., 2013. Adaptation of lift forces in object manipulation through action observation. *Experimental brain research*, 228(2), pp.221–234.

Richard, H. et al., 2020. Modeling shared responses in neuroimaging studies through multiview ica. *arXiv preprint arXiv:2006.06635*.

Ridderinkhof, K.R. et al., 2004. The role of the medial frontal cortex in cognitive control. *science*, 306(5695), pp.443–447.

Ronsse, R. & Sternad, D., 2010. Bouncing between model and data: Stability, passivity, and optimality in hybrid dynamics. *Journal of motor behavior*, 42(6), pp.389–399.

Satterthwaite, T.D. et al., 2013. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64(1), pp.240–256. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1053811912008609.

Satterthwaite, T.D., Xia, C.H. & Bassett, D.S., 2018. Personalized neuroscience: Common and individual-specific features in functional brain networks. *Neuron*, 98(2), pp.243–245.

Saygin, Z.M. et al., 2012. Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature neuroscience*, 15(2), pp.321–327.

Schaal, S. & Atkeson, C.G., 1993. Open loop stable control strategies for robot juggling. In *[1993] proceedings ieee international conference on robotics and automation*. IEEE, pp. 913–918.

Schaal, S., Atkeson, C.G. & Sternad, D., 1996. One-handed juggling: A dynamical approach to a rhythmic movement task. *Journal of Motor Behavior*, 28(2), pp.165–183.

Scholz, J. et al., 2009. Training induces changes in white-matter architecture. *Nature neuroscience*, 12(11), pp.1370–1371.

Schölvinck, M.L., Howarth, C. & Attwell, D., 2008. The cortical energy needed for conscious perception. *Neuroimage*, 40(4), pp.1460–1468.

Schultz, D.H. & Cole, M.W., 2016. Higher intelligence is associated with less task-related brain network reconfiguration. *Journal of Neuroscience*, 36(33), pp.8551–8561.

Seriès, P., Latham, P.E. & Pouget, A., 2004. Tuning curve sharpening for orientation selectivity: Coding efficiency and the impact of correlations. *Nature neuroscience*, 7(10), pp.1129–1135.

Shadmehr, R. & Holcomb, H.H., 1997. Neural correlates of motor memory consolidation. *Science*, 277(5327), pp.821–825.

Shadmehr, R. & Mussa-Ivaldi, F.A., 1994. Adaptive representation of dynamics during learning of a motor task. *Journal of neuroscience*, 14(5), pp.3208–3224.

Shadmehr, R., Smith, M.A. & Krakauer, J.W., 2010. Error correction, sensory prediction, and adaptation in motor control. *Annual review of neuroscience*, 33, pp.89–108.

Shannon, B.J. et al., 2016. Brain aerobic glycolysis and motor adaptation learning. *Proceedings of the National Academy of Sciences*, 113(26), pp.E3782–E3791.

Shen, H.H., 2015. Core concept: Resting-state connectivity. *Proceedings of the National Academy of Sciences*, 112(46), pp.14115–14116.

Shulman, G.L. et al., 1997. Common blood flow changes across visual tasks: I. Increases in subcortical structures and cerebellum but not in nonvisual cortex. *Journal of cognitive neuroscience*, 9(5), pp.624–647.

Sidarta, A. et al., 2016. Somatic and Reinforcement-Based Plasticity in the Initial Stages of Human Motor Learning. *The Journal of Neuroscience*.

Siegler, I., Bazile, C. & Warren, W., 2013. Mixed control for perception and action: Timing and error correction in rhythmic ball-bouncing. *Experimental brain research*, 226(4), pp.603–615.

Sigman, M. et al., 2005. Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron*, 46(5), pp.823–835.

Smith, M.A. & Shadmehr, R., 2005. Intact ability to learn internal models of arm dynamics in huntington's disease but not cerebellar degeneration. *Journal of neurophysiology*, 93(5), pp.2809–2821.

Smith, S.M. & Brady, J.M., 1997. SUSAN—a new approach to low level image processing. *International journal of computer vision*, 23(1), pp.45–78.

Smith, S.M. et al., 2009. Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, 106(31), pp.13040–13045.

Smith, S.M. & Nichols, T.E., 2009. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, 44(1), pp.83–98.

Snoddy, G.S., 1926. Learning and stability: A psychophysiological analysis of a case of motor learning with clinical applications. *Journal of Applied Psychology*, 10(1), p.1.

Sochat, V.V. et al., 2015. Effects of thresholding on correlation-based image similarity metrics. *Frontiers in neuroscience*, 9, p.418.

Steel, A. et al., 2019. Finding the baby in the bath water–evidence for task-specific changes in resting state functional connectivity evoked by training. *NeuroImage*, 188, pp.524–538.

Stelzer, J. et al., 2014. Deficient approaches to human neuroimaging. *Frontiers in Human Neuroscience*, 8, p.462.

Sternad, D., 2017. Human control of interactions with objects–variability, stability and predictability. In *Geometric and numerical foundations of movements*. Springer, pp. 301–335.

Sternad, D. et al., 2000. Dynamics of a bouncing ball in human performance. *Physical Review E*, 63(1), p.011902.

Strappini, F. et al., 2019. Resting-state activity in high-order visual areas as a window into natural human brain activations. *Cerebral Cortex*, 29(9), pp.3618–3635.

Sutton, R.S., Barto, A.G. & others, 1998. *Introduction to reinforcement learning*, MIT press Cambridge.

Tambini, A., Ketz, N. & Davachi, L., 2010. Enhanced brain correlations during rest are related to memory for recent experiences. *Neuron*, 65(2), pp.280–290.

Taubert, M. et al., 2010. Dynamic properties of human brain structure: Learning-related changes in cortical areas and associated fiber connections. *Journal of Neuroscience*, 30(35), pp.11670–11677.

Taubert, M. et al., 2011. Long-term effects of motor training on resting-state networks and underlying brain structure. *NeuroImage*.

Tavor, I. et al., 2020. Short-term plasticity following motor sequence learning revealed by diffusion magnetic resonance imaging. *Human brain mapping*, 41(2), pp.442–452.

Tavor, I. et al., 2016. Task-free mri predicts individual differences in brain activity during task performance. *Science*, 352(6282), pp.216–220.

Telesford, Q.K., Ashourvan, A. & Wymbs, N.F., 2017. Cohesive network reconfiguration accompanies extended training. *Human Brain Mapping*.

Thomas, A.G. et al., 2009. Functional but not structural changes associated with learning: An exploration of longitudinal voxel-based morphometry (vbm). *Neuroimage*, 48(1), pp.117–125.

Tobyne, S.M. et al., 2017. Sensory-biased attention networks in human lateral frontal cortex revealed by intrinsic functional connectivity. *Neuroimage*, 162, pp.362–372.

Tobyne, S.M. et al., 2018. Prediction of individualized task activation in sensory modality-selective frontal cortex with "connectome fingerprinting". *NeuroImage*, 183, pp.173–185.

Todorov, E. & Jordan, M.I., 2002. Optimal feedback control as a theory of motor coordination. *Nature neuroscience*, 5(11), pp.1226–1235.

Tomasi, D., Wang, G.-J. & Volkow, N.D., 2013. Energetic cost of brain functional connectivity. *Proceedings of the National Academy of Sciences*, 110(33), pp.13642–13647.

Toni, I. et al., 1998. The time course of changes during motor sequence learning: A whole-brain fMRI study. *Neuroimage*, 8(1), pp.50–61.

Toro, R., Fox, P.T. & Paus, T., 2008. Functional coactivation map of the human brain. *Cerebral cortex*, 18(11), pp.2553–2559.

Tufillaro, N. & Albano, A., 1986. Chaotic dynamics of a bouncing ball. *American Journal of Physics*, 54(10), pp.939–944.

Tufillaro, N.B., Abbott, T. & Reilly, J., 1992. *An experimental approach to nonlinear dynamics and chaos*, Addison-Wesley Redwood City, CA.

Turek, J.S. et al., 2017. A semi-supervised method for multi-subject fMRI functional alignment. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE, pp. 1098–1102.

Turner, B.O. et al., 2018. Small sample sizes reduce the replicability of task-based fMRI studies. *Communications biology*, 1(1), p.62.

Turvey, M.T. & Fonseca, S., 2009. Nature of motor control: Perspectives and issues. In *Progress in motor control*. Springer, pp. 93–123.

Tustison, N.J. et al., 2010. N4ITK: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6), pp.1310–1320.

Uddin, L.Q., 2020. Bring the noise: Reconceptualizing spontaneous neural activity. *Trends in Cognitive Sciences*.

Uddin, L.Q. et al., 2009. Functional connectivity of default mode network components: Correlation, anti-correlation, and causality. *Human brain mapping*, 30(2), pp.625–637.

Uddin, L.Q., Yeo, B.T. & Spreng, R.N., 2019. Towards a universal taxonomy of macro-scale functional human brain networks. *Brain topography*, pp.1–17.

Ungerleider, L.G., Doyon, J. & Karni, A., 2002. Imaging brain plasticity during motor skill learning. *Neurobiology of learning and memory*, 78(3), pp.553–564.

Urner, M. et al., 2013. Early visual learning induces long-lasting connectivity changes during rest in the human brain. *NeuroImage*, 77, pp.148–156.

Vahdat, S. et al., 2011. Functionally specific changes in resting-state sensorimotor networks after motor learning. *Journal of Neuroscience*, 31(47), pp.16907–16915.

Vahdat, S., Darainy, M. & Ostry, D.J., 2014. Structure of plasticity in human sensory and motor networks due to perceptual learning. *Journal of Neuroscience*, 34(7), pp.2451–2463.

Van Den Heuvel, M.P. et al., 2009. Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience*, 29(23), pp.7619–7624.

Van Dijk, K.R. et al., 2009. Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization. *Journal of neurophysiology*, 103(1), pp.297–321.

Van Essen, D.C. et al., 2013. The wu-minn human connectome project: An overview. *Neuroimage*, 80, pp.62–79.

Vanwinckelen, G. & Blockeel, H., 2012. On estimating model accuracy with repeated cross-validation. In *Benelearn 2012: Proceedings of the 21st belgian-dutch conference on machine learning*. pp. 39–44.

Varoquaux, G., 2018. Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage*, 180, pp.68–77.

Varoquaux, G. et al., 2011. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Biennial international conference on information processing in medical imaging*. Springer, pp. 562–573.

Varoquaux, G. et al., 2017. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, 145, pp.166–179.

Varoquaux, G. et al., 2010. A group model for stable multi-subject ica on fMRI datasets. *Neuroimage*, 51(1), pp.288–299.

Vatansever, D., Menon, D.K. & Stamatakis, E.A., 2017. Default mode contributions to automated information processing. *Proceedings of the National Academy of Sciences*, 114(48), pp.12821–12826.

Viessmann, O. et al., 2019. Dependence of resting-state fMRI fluctuation amplitudes on cerebral cortical orientation relative to the direction of b0 and anatomical axes. *NeuroImage*, 196, pp.337–350.

Vincent, J.L. et al., 2007. Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, 447(7140), pp.83–86.

Von Helmholtz, H., 1867. *Handbuch der physiologischen optik: Mit 213 in den text eingedruckten holzschnitten und 11 tafeln*, Voss.

Warren, W.H., 2006. The dynamics of perception and action. *Psychological review*, 113(2), p.358.

Wehbe, L. et al., 2015. Regularized brain reading with shrinkage and smoothing. *The Annals of Applied Statistics*, 9(4), pp.1997–2022.

Welvaert, M. & Rosseel, Y., 2014. A review of fMRI simulation studies. *PloS one*, 9(7), p.e101953.

Wickelgren, W.A., 1977. Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica*, 41(1), pp.67–85.

Wiestler, T. & Diedrichsen, J., 2013. Skill learning strengthens cortical representations of motor sequences. *Elife*, 2, p.e00801.

Wilf, M. et al., 2017. Spontaneously emerging patterns in human visual cortex reflect responses to naturalistic sensory stimuli. *Cerebral cortex*, 27(1), pp.750–763.

Wilson, R.C. & Niv, Y., 2015. Is model fitting necessary for model-based fMRI? *PLoS computational biology*, 11(6), p.e1004237.

Winkler, A.M. et al., 2014. Permutation inference for the general linear model. *Neuroimage*, 92, pp.381–397.

Wollman, I. et al., 2018. Neural network retuning and neural predictors of learning success associated with cello training. *Proceedings of the National Academy of Sciences*, 115(26), pp.E6056–E6064.

Wolpert, D.M., 1997. Computational approaches to motor control. *Trends in cognitive sciences*, 1(6), pp.209–216.

Wolpert, D.M., Diedrichsen, J. & Flanagan, J.R., 2011. Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(12), pp.739–751.

Woolrich, M.W. et al., 2001. Temporal autocorrelation in univariate linear modeling of fmri data. *Neuroimage*, 14(6), pp.1370–1386.

Wulf, G. & Shea, C.H., 2002. Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic bulletin & review*, 9(2), pp.185–211.

Xu, J. et al., 2013. Evaluation of slice accelerations using multiband echo planar imaging at 3 t. *Neuroimage*, 83, pp.991–1001.

Yang, C.S., Cowan, N.J. & Haith, A.M., 2020. De novo learning versus adaptation of continuous control in a manual tracking task. *bioRxiv*.

Zatorre, R.J., Fields, R.D. & Johansen-Berg, H., 2012. Plasticity in gray and white: Neuroimaging changes in brain structure during learning. *Nature neuroscience*, 15(4), pp.528–536.

Zhang, H., Chen, P.-H. & Ramadge, P., 2018. Transfer learning on fMRI datasets. In *International conference on artificial intelligence and statistics*. pp. 595–603.

Zhang, Y., Brady, M. & Smith, S., 2001. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), pp.45–57.