7-20-2017 12:00 AM

Evidence in Neuroimaging: Towards a Philosophy of Data Analysis

Jessey Wright, *The University of Western Ontario*

Supervisor: Dr. Jacqueline Sullivan, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Philosophy

Abstract

Neuroimaging technology is the most widely used tool to study human cognition. While originally a promising tool for mapping the content of cognitive theories onto the structures of the brain, recently developed tools for the analysis, handling and sharing of data have changed the theoretical landscape of cognitive neuroscience. Even with these advancements philosophical analyses of evidence in neuroimaging remain skeptical of the promise of neuroimaging technology. These views often treat the analysis techniques used to make sense of data produced in a neuroimaging experiment as one, attributing the inferential limitations of analysis pipelines to the technology as a whole. Situated against the neuroscientists' own critical assessment of their methods and the limitations of those methods, this skepticism appears based on a misunderstanding of the role data analysis techniques play in neuroimaging research. My project picks up here, examining how data analysis techniques, such as pattern classification analysis, are used to assess the evidential value of neuroimaging data. The project takes the form of three papers. In the first I identify the use of multiple data analysis techniques as an important aspect of the data interpretation process that is overlooked by critics. In the second I develop an account of inferences in neuroimaging research that is sensitive to this use of data analysis techniques, arguing that interpreting neuroimaging data is a process of isolating and explaining a variety of data patterns. In the third I argue that the development and uptake of new techniques for analyzing data must be accompanied by changes in research practices and standards of evidence if they are to promote knowledge generation. My approach to this work is both traditionally philosophical, insofar as it involves reading and analyzing the work of philosophers and neuroscientists, and embedded insofar as most of the research was conducted while attending lab meetings and participating in the work of those scientists whose work is the object of my research.

Keywords

Philosophy of Science, Philosophy of Neuroscience, Epistemology of Experiment, Neuroimaging, Data, Data Analysis, Pattern Classification Analysis, Multivariate Pattern Analysis, Explanation.

Acknowledgements

I am grateful to the support, encouragement, advice, and guidance of all of the mentors and peers who have been part of the incredible process of which this document is the final product. I could not hope to name every individual and community that played a role. There are a number, however, who made this project possible.

I am deeply grateful for the support, mentorship, guidance and, above all, patience of supervisor, Dr. Jacqueline Sullivan. It was Jackie's class on the philosophy of neuroscience that inspired my project, and her supervision and direction that saw me develop the skills and knowledge necessary to pursue those ideas. I am also grateful for the feedback and support of other participants in the lab associates program in the Rotman Institute of Philosophy who shared in the journey of learning and engaging with the practice and practitioners of cognitive neuroscience at the Brain and Mind Institute. I am especially grateful for the support and feedback provided by Robert Foley, Frédéric Banville and Daniel Booth at almost every stage of this project. I am grateful for those who made the Lab Associates program possible, including Jacqueline Sullivan, Chris Viger and Robert Foley for their hard work creating, maintaining and facilitating the program. I owe much to the community of neuroscientists at the Brain and Mind Institute for their generosity in openly welcoming philosophers, and Stefan Köhler for welcoming me into his lab. From the lab, I am especially thankful for the conversations I had with Chris Martin, Edward O'Neil, Anna Blumenthal, and Jordan Dekraker. Additional thanks to all other members of the Köhler Memory Lab at the Brain and Mind Institute for productive and insightful discussions on their research and methods, and the open and constructive environment of the lab and lab meetings.

I am deeply grateful and thankful for my partner, wife and closest friend, Helana Hope. Without her continual support, encouragement, and understanding, I could not have started this journey, nor could I have finished this project.

Funding for parts of this research was provided by the Social Sciences and Humanities Research Council of Canada, and the Rotman Institute of Philosophy.

Table of Contents

Chapter 1

1 Evidence in Neuroimaging

Neuroimaging technology is the most widely used technology in cognitive neuroscience to study the human brain. Neuroimaging experiments involve measurements of blood oxygenation levels and participant behaviour, which are used as evidence for claims about the relationship between cognitive processes and neural activity. Techniques of data manipulation and analysis are what, in practice, bridge the gap between the objects of measurement and the phenomena that neuroscientists are interested in learning about. To motivate the three papers that follow and situate them as addressing a common challenge, in this introductory chapter I examine and contrast the philosophical and neuroscientific views on the promise and perils of data analysis techniques as used to interpret neuroimaging data.

1.1 Introduction

Neuroimaging data are large, complex and laden with uncertainties. A single scanning session, which a neuroimaging experiment includes at least twelve of, can produce over 20,000 data points, making neuroimaging data 'big' by most measures. The data are complex as they include measurements of blood oxygenation, behaviour and brain structure. Uncertainties arise from the fact that the details about the relationship between the data points and phenomena of interest are not fully known. The strategy neuroscientists use to bring this data to bear as evidence on claims about the phenomena they are interested in involves the application of methods of data analysis and manipulation. The broad aim of this project is to develop an account of how methods of data analysis and manipulation are used to overcome challenges with interpreting neuroimaging data. This is done in the context of ongoing debate between skeptics of the technology and critical advocates as contrasted with the methodological debates that occur within the neuroscientific literature.

Functional magnetic resonance imaging (fMRI) is the most widely used methodology that modern neuroscience has at its disposal to investigate the healthy human brain in

action; fMRI data are used to support claims about the involvement of particular brain regions in performing specific cognitive functions, and in representing different types of information. The technology provides insight into variations in neural activity via measurements of blood oxygenation levels in the brain — called the blood oxygenation level dependant, or BOLD, signal. In human neuroimaging research, BOLD measurements are collected while participants perform tasks in the scanner. Participant behaviour and facts about the task parameters are used to relate variations in BOLD activity to cognitive processes, capacities and states.

In the introduction to a volume of reflections on the then 20 year history of fMRI, Peter Bandettini explains that "[e]ven though the underlying relationships between changes in brain activation and changes in BOLD contrast-weighted MRI signal are still debated, it's clear that the method has proved itself more robust, reliable, and information-rich than most originally anticipated" (2012, p. 576). The surprise is in part due to, as Bandettini notes, uncertainty about the relationship between the measured data points and phenomena the data are used to make claims about. Functional scanning protocols measure changes in blood oxygenation, which is, at best, a proxy for cognitively relevant neural activity. Even though the measurements are of causal factors indirectly related to the neural and cognitive systems neuroscientists are interested in, neuroimaging data has been put to some surprising uses in the field. Relatively recently neuroimaging data has come to be viewed as valuable for pursuing a deeper understanding of how information is processed and represented in the brain. Kenneth Norman and colleagues, for example, note that "[f]unctional MRI (fMRI) is a powerful tool for addressing questions…" such as "… what information is represented in different brain structures; and how is that information transformed at different stages of processing?" (Norman et al 2006, p. 424).

The indirect and uncertain relationship between the objects of measurement and phenomena of interest makes it prudent to ask: Is the BOLD signal really as valuable for the study of human cognition as the current practice in neuroscience takes it to be, or are these research projects reaching well beyond the evidence? The philosophical literature on neuroimaging research tends to favour a skeptical answer, stocked as it is with regularly rehearsed arguments that theories and claims neuroscientists often infer on the

basis of neuroimaging data are not justified by the available evidence (van Orden and Paap 1997; Uttal 2001; Hardcastle and Stewart 2002; Aktunç 2014; Ritchie, Kaplan and Klein forthcoming). In addition to their skeptical conclusions and apprehensive stance towards neuroimaging research, the arguments offered by critics of the technology often include a detailed analysis of the logic implicit in the methods of data analysis applied in the interpretation of neuroimaging data. As it was the dominant method of analysis in the early days of neuroimaging research, skeptics have tended to emphasize the problems with subtractive methods of analysis (beginning with van Orden and Paap 1997). While new analysis methods have entered the field in the last decade and a half, the focus on subtraction has continued in the skeptical literature (e.g., Aktunç 2014). That is, at least, until very recently (e.g., Ritchie, Kaplan and Klein forthcoming).

It is no accident that the primary point of contact between skeptics and the practice of neuroimaging research are the methods of data analysis that dominate the field. Methods of data analysis, like subtraction or the newer machine-learning inspired methods that are rapidly taking over, are a central part of neuroimaging research methods. Indeed, the progress alluded to above has, at least in part, been driven by the development of new techniques for analyzing and manipulating data. Examples of innovations in the methods of analysis and data manipulation driving progress include the discovery of the default mode network, which is a collection of brain regions that have been shown to reliably co-activate when subjects are required to 'do nothing' in the scanner (e.g., Gusnard & Raichle 2001), and research on the representational character of brain activity patterns (e.g., Tong and Pratte 2012). The discovery of the default mode network was partly based on manipulating data differently during their analysis. That is, "… researchers began routinely noticing brain regions more active in the passive control conditions than the active target tasks" (Bucker, Andrews-Hanna and Schater 2008, p. 3), and so began to see control conditions as potentially reflecting the influence of a shared phenomenon. This became what is now known as the default mode network. With respect to research on representations, the critical development seems to have been the uptake of machine learning methods of data analysis that bring "... fMRI investigation closer to investigating the codes for how functions are represented in neural population responses…" (Haxby 2010, p. 56). In both of these cases, changes in the way neuroimaging data was

manipulated and analyzed were critical for making discoveries, or changing scientist's views on what claims the available data are relevant for learning about.

Given the central role of methods of data analysis, it should not be surprising that neuroscientists frequently engage in discussion and debate over the uses and limitations of techniques like subtraction and the machine learning methods behind research on how information is represented in the brain. As it turns out (and as I discuss in more detail in the next section), many of the assumptions identified as problematic by skeptics parallel the limitations of the analysis methods that are openly discussed and debated in the scientific literature. What differs is the response to these challenges. Where skeptics argue for tempering conclusions and limiting the scope of neuroimaging research, neuroscientists forge ahead. The different conclusions may either be due to skeptics overlooking some important feature of the research that blunts the force of their critiques, or neuroscientists failing to grasp the significance of these challenges. The responses to skeptics available in the philosophical literature, however, tilts the scales in favour of the neuroscientists. Critical advocates of the research tend to resist skeptical arguments by identifying how the skeptic in question has overlooked or oversimplified an epistemically relevant aspect of the practice. This is not to say that research practices in neuroimaging research are flawless. Only that it seems, from this cursory and top-down viewpoint, that the positions offered by critics miss something about the role of data analysis techniques in neuroimaging research. This raises the question that is central to my project: what contribution do data analysis techniques make to the inferential practices operative in neuroimaging research?

As the skeptical perspective appears to consistently miss important aspects of the research practices neuroscientists engage in, I have explicitly adopted an approach to conceptualizing and analyzing research practices that is distinct from the form of features of the skeptics' arguments. Instead of examining the structure of inferences and articulating the conditions that must obtain for those inferences to be warranted, then assessing if those conditions do in fact obtain in practice, I examine the procedures neuroscientists engage in to interpret and make sense of neuroimaging data. In doing so, my aim is to uncover the various factors that contribute to the perceived value of

neuroimaging data that ultimately result in inferences to claims. I treat assessments of the evidential value of data, or the process of data interpretation, as the primary object of my analyses, and not the form of the inference that results from that process. I have pursued this project piecemeal, writing three research papers each engaging with a question related to judgements of the evidential significance of neuroimaging data. In the first I articulate how skeptical arguments overlook certain uses of data analysis techniques by virtue of isolating the analysis process from the broader research context. The second paper argues that data analysis techniques support data interpretation through the isolation of data patterns that can be explained by appeal to claims about the phenomena of interest. In the third paper I argue that the use and uptake of large scale databases and meta-analysis tools must be accompanied by the development and uptake of research practices appropriate for interpreting the resulting data sets if it is to be epistemically advantageous.

My research for this project has involved regular interactions with neuroscientists via the lab associates program available through a mutually beneficial arrangement between the Rotman Institute of Philosophy and the Brain and Mind Institute at the University of Western Ontario. This has afforded me opportunities to collaborate on research projects (e.g., Martin et al 2015), attend lab meetings, and regularly interact with members of the brain and mind institute in general, and the Köhler memory lab in particular. These experiences underpin many of my arguments and views that follow. As casual and regular interactions are not inherently convincing or reliable data points, I support the insights gained from my experience in the lab with textual analyses of research papers published in neuroscience journals.

With the remainder of this introductory chapter I provide the background and further motivation for this project by examining the philosophical and neuroscientific perspectives on data analysis in neuroimaging research. The next section presents the view from neuroscience. I provide an overview of the significance and promise of data analysis techniques, and some of the concerns that neuroscientists have raised with respect to the current state of the field. In the third section I present the view from philosophy. There, I briefly outline two philosophical debates that pertain to inferences in

neuroimaging research. On one hand, there is the ongoing debate about the claims neuroimaging data can and cannot be used to infer, and on the other is Jim Bogen's argument that counterfactual (Woodward 2000) and error-statistical (Mayo 1996) accounts of experimental evidence fail to identify what makes neuroimaging data epistemically valuable (2001; 2002). Contrasting the views from neuroscience and philosophy raises a number of questions about the significance of data analysis for the interpretation of neuroimaging data that are taken up in the papers that follow. In the fourth section I present an overview of those papers, and roughly situate their individual contributions within the larger aims of this project.

1.2 The View from Neuroscience

Neuroimaging data are used by neuroscientists as evidence for claims about the relationship between cognitively relevant brain activity and cognitive processes, states, or capacities. While experimental design plays an important role in establishing the evidential value of neuroimaging data, data analysis techniques act as a bridge between the data that is produced in an experiment and the claims neuroscientists take those data to be evidence for. New methods for the analysis of neuroimaging data are often developed in the pursuit of a particular question or hypothesis, and some of these go on to take on a life of their own as they are refined and see uptake throughout the broader community. This process is only possible because neuroimaging data itself is rich enough to be useful for addressing research questions beyond those it is produced to investigate.

The phenomena cognitive neuroscientists use neuroimaging technology to investigate requires information about the structure, activity and connectivity of the brain to be brought together with information about cognitive processes. Each of these experimental targets is accessed by different modes of measurement within a neuroimaging experiment.

Experiments using fMRI involve placing a human subject in a magnetic resonance imaging scanner (MRI) while they perform a cognitive task, such as attending to a moving pattern of dots (Liu et al 2011), or deciding whether or not an image is one they saw in a previous part of the experiment (Martin et al 2013). While the participant

performs the task, the scanner measures changes in brain activity. The BOLD signal data captured by fMRI provides information about the activity of the brain, MRI scanning protocols provide structural information, while the cognitive states of subjects are probed through the use of carefully designed cognitive tasks. Each of these produce distinct data sets that are used together to make inferences in neuroimaging research.

There are many ways to integrate, analyze and classify these data sets. Since each decision made in the process of integrating, manipulating and analyzing data is informed by the aims of the scientists, it is in principle possible to use neuroimaging data to investigate phenomena that they were not originally produced to investigate by analyzing them in different ways. Indeed, advocates of open databases often argue for data sharing by emphasizing the need to capitalize on this potential. With respect to neuroimaging data, the argument is that "[i]f such data can be archived, indexed with accompanying meta-data, and combined, there is an enormous opportunity to obtain deep insights into the workings of the brain and mind" (p. 678) since "... there are often dimensions of the data that are not fully explored or even recognized by the researchers obtaining it…" (Van Horn and Gazzaniga 2013, p. 678).

An example of the potential for making new discoveries through the reuse of neuroimaging data can be found in the discovery of the default mode network, which created a sub-field of research on resting state fMRI, and is an important component of the Human Connectome Project (Smith et al 2013). Early fMRI research treated task-free conditions as a baseline, where task-free conditions require the participant to 'do nothing', let their 'mind wander', or otherwise remain still in the scanner without executing a particular task. These task-free conditions were typically used as a baseline contrast for a task-based condition in research aiming to isolate task-relevant brain activity through subtraction analyses. The discovery of networks of brain activity persistent across subjects in resting state was made possible by a combination of a choice in the analysis of neuroimaging data — that is to look at the 'task free' data — and the availability of a large volume of useable data. Analyses of task-free data revealed a collection of brain regions that are consistently active across subjects. Those patterns are

now thought to reflect the activity of a ‘default mode network’ (Gusnard & Raichle 2001; Greicius et al 2003; Morcom and Fletcher 2007).

While an example of how new ways of looking at neuroimaging data can lead to discoveries, the story of default mode network isn’t driven by the development of a new technique for the analysis of data. The recent trend towards research aimed at understanding representations in the brain, however, provides a clearer example of this. Traditionally neuroimaging data was used to identify the regions of the brain associated with cognitive tasks, a theoretical project sometimes referred to as the localization of cognitive functions. This was often done by analyzing fMRI data using subtraction analysis. Subtraction involves taking the difference between two BOLD signal data sets associated with distinct task conditions and attributing the difference in measured BOLD signal to the cognitive difference between the tasks.

More recently neuroimaging data has been used to identify the informational content of brain activity and to investigate where and how representations of stimuli are contained and processed in the brain (Tong and Pratte 2012). This relatively new use of neuroimaging data has been driven by the development of new techniques for analyzing and interpreting it. These include machine learning tools such as pattern classification analysis (Haxby 2010), and representational similarity analysis (Kriegeskorte and Kievit 2013). Advocates of these techniques argue that they can be used to answer new questions. For instance, an introduction to pattern classification techniques identifies three new questions that they can be used to address. They are: Is there information about a variable of interest? Where is the information? And, how is that information encoded? (Pereira, Mitchell and Botvinick 2009, p. S208). Some even go so far as to argue that "[i]n addition to allowing us to sensitively detect and track cognitive states, MVPA methods can be used to characterize how these cognitive states are represented in the brain" (Norman et al 2006, p. 425).

The paper often cited as pioneering these techniques approached data analysis with a technique inspired by machine learning to discriminate between three hypotheses about the functional architecture of the ventral visual pathway (Haxby et al 2001). Of the three

hypotheses, two proposed modular architectures in which distinct parts of the region are specialized for processing or representing information about a particular category of object or performing a particular process. The third proposed that the "representations of faces and different categories of objects are widely distributed and overlapping" (p. 2425). The new method of data analysis used in the study, which involved determining if variations in the BOLD signal could be used to predict the object category (face, house, cat, etc) of the stimulus correlated with it, was chosen for its capacity to discriminate between modular and distributed processing hypotheses.

This data analysis technique was introduced to the field because researchers saw in it the potential to provide information relevant to the assessment of a claim that the functional architecture of a brain region is distributed. This marks an important change in the epistemic landscape of neuroimaging research. Consider, a common criticism of localization research that uses neuroimaging data is that the analysis procedure often used, that is subtraction, assumes that cognitive processing is not distributed across the cortex, or even within a larger region (e.g., Uttal 2001; Hardcastle and Stewart 2002). Localization hypotheses, critics argue, are assumed by virtue of the analysis technique used to confirm them. If the analysis method Haxby and colleagues used can in fact discriminate between modular and distributed processing hypotheses of brain architecture, then the availability of these new data analysis method renders this line of skepticism obsolete.

Haxby and colleagues' result was important as it set a new threshold of evidence for modular theories of functional architecture. They demonstrated that it is possible, and necessary, to not only show preferential activation but also to show that the regional activity carries information about the relevant stimuli (as noted in Kanwisher 2017). Separately from its impact on theories of ventral visual stream architecture, it acted as a proof of concept for a new approach to the analysis of neuroimaging data. The little machine learning inspired methods presented in their paper has since grown into a rich and diverse collection of analysis techniques broadly referred to as multivariate pattern analysis (MVPA).

The BOLD signal is now, when analyzed with MVPA techniques, used to pursue a broader range of theoretical aims than merely assigning cognitive functions to discrete regions of the brain, as was the standard contribution of the technology in its early years. Pattern classification techniques, for instance, have been used to show that variations in BOLD activity can predict the features of a stimuli a participant is attending to (Kamitani and Tong 2005), to test whether the perception of an object and the act of imagining that object share a representational profile (Reddy, Tsuchiya and Serre 2010), to address a number of confounds and challenges in the study of consciousness (Sandberg, Andersen and Overgaard 2014), and has been identified as a solution to the problem of 'reverse inference' in neuroimaging research (Poldrack 2011). The last of these provides another example of innovations in data analysis changing the epistemic landscape of cognitive neuroscience.

Reverse inference refers to the use of brain activity data to ascribe a cognitive state to a subject, such as inferring that a subject is experiencing fear on the basis of an observation of activation in their amygdala (an area commonly associated with fear). Reverse inferences are common in neuroimaging papers, and while often informal there are studies for which reverse inferences are a central result. The problem is that neuroimaging experiments involve manipulations of behaviour and measurements of correlated changes in brain activity, while reverse inferences start from brain activity and move to ascriptions of cognitive states. This has been identified by neuroscientists as committing the logical fallacy of 'affirming the consequent' (Poldrack 2006, p. 2). Furthermore, meta-analysis evidence provided in the same paper suggests that reverse inferences are unreliable as 'best explanations' for brain activity patterns observed in an experiment because any given region of the brain is implicated in a wide range of cognitive processes and so the activity could be reflecting any one of those (p. 4-5). The same neuroscientist who classified reverse inferences as a fallacy, has, with the uptake of MVPA methods, recently changed his view. He has argued that pattern classification techniques in particular "… provide a formal means to implement reverse inference" (2011, p. 4). One of the problems with reverse inference is that the evidence available in a neuroimaging experiment cannot discriminate between two competing reverse inference claims. Pattern classification techniques are used to evaluate if patterns and

variations within one set of variables can be used to predict, or identify, the values of correlated variables. This allows neuroscientists to directly evaluate whether or not variations in brain activity are predictive, or diagnostic, of the engagement of specific tasks and cognitive states, an evaluation that simply was not possible prior to the development of these methods. Treating this change in perspective at face value, the problem with reverse inferences is resolved not by building better measurement devices, designing better experiments or considering theoretical constraints arising from other areas of neuroscience. The problem with reverse inferences is solved by analyzing imaging data with an analysis technique suitable for evaluating reverse inference claims.

How data analysis techniques are used, including the decisions made in their implementation as well as the form of the results scientists focus on, impact the phenomena neuroscientists recognize data as relevant for learning about. Pattern classification analysis is one of the more popular MVPA techniques. It involves training a machine learning classifier to predict task conditions (e.g., ‘face’, ‘chair’, ‘building’), or behavioural responses (e.g., ‘remembered’, ‘forgotten’) based on variations in the BOLD signal that are correlated with those conditions or responses, then testing it on novel data and evaluating its accuracy. Investigators typically focus on the accuracy of the classifier at performing a primary classification task, such as identifying which of two patterns of moving dots a subject is paying attention to (as in Liu et al 2011). Classifier accuracy is used to support claims about the information carried by patterns of brain activity. If a classifier is able to discriminate between different attentional conditions, one might conclude that there is information in the brain activity patterns originating in the region of interest relevant for making such a discrimination. Determining whether or not a classifier can accurately classify is not always the only dimension of the analysis process that is relevant. Classification failures can sometimes be important and informative when they are found to correlate with behavioural errors, as they allow investigators to begin to draw closer relations between the information contained in the brain activity and information that the participant is acting on (Walther 2012, for instance, proposes a mathematical method for doing so). Choices about the data the classifier is using can also be informative to consider when evaluating certain hypotheses. Tambini and Davachi, for instance, test and train a classifier using data from different time points after the task of

interest to evaluate whether or not activity patterns permitting classification persist in the same way our actual memories of events do (2013). Comparing different analysis results is often useful for better understanding what it means for a classifier to be able to accurately identify the task conditions correlated with BOLD measurements. For example, classifier accuracy is often contrasted with subtraction results to assist in determining if information in the activity pattern requires the full multi-dimensional representation that classifiers leverage, or if average or mean BOLD activity is sufficient for consistently linking the brain activity to the task condition (Kohler et al 2013; Coutanche 2013). These examples show that the evidence provided by a data analysis process depends in part on the decisions investigators make when implementing it. These include decisions about how to implement an analysis technique, which variables and results to focus on, and the supplementary methods investigators choose to use.

The development and uptake of new analysis techniques is treated by many neuroscientists with cautious optimism. In light of the kinds of developments discussed above, there are three concerns raised as barriers to the promise of progress that innovations in data analysis bring with them. Those concerns are: (1) the impact of an increase in analytic flexibility on false positive rates, (2) technical analytic skills overriding practical intuitions, and (3) the risks of replication failures.

Analytic flexibility refers to the "… range of analysis outcomes across different acceptable analysis methods" (Carp 2012, p. 1), and has been associated with an increase risk for inflating the rate of false positives within a research domain (Ioannidis 2005). Joshua Carp examined the analytic flexibility in fMRI analysis pipelines, where a pipeline is a series of data manipulations that produce an interpretable result (2012). Variations in the order of manipulations, including those used to reduce error and correct for artifacts, as well as variations in the statistical techniques themselves constitute distinct pipelines. Carp found that, while some outcomes were consistent across a wide range of pipelines, other results varied substantially. This flexibility means that, "… a motivated researcher determined to find significant activation in practically any brain region will very likely succeed…" (p. 12). Flexibility isn't only worrisome in the case of a 'motivated researcher', but creates a situation in which a well-intentioned researcher

may infer claims that would not have appeared to be supported by the data had they made different decisions in their analysis procedure. Andrew Gelman and Eric Loken, for instance, argue that "… researchers can perform a reasonable analysis given their assumptions and their data, but had the data turned out differently, they could have done other analyses that were just as reasonable in those circumstances" (Gelman and Loken 2013, p. 1). These results appear to have all the significance of well-conducted statistical tests, and yet, due to the degrees of freedom in the analysis that are not apparent to the researchers conducting it, can have the same epistemic standing as the results of p-hacking or the work of a 'motivated researcher' seeking any 'publishable result of statistical significance'.

Analytic flexibility and its impact on false positive rates is a significant concern for neuroimaging research when considering only the variations in standard pipelines. Introducing new methods for data analysis, such as those briefly discussed above, creates more opportunity for this kind of inferential error to occur. This worry is amplified by the fact that these new techniques are more sophisticated than their predecessors. Eve Marder, reflecting on the history and future of neuroscience, notes that, in this era of high volumes of data and complex techniques for analyzing it, "… new findings will depend on data analyses that are highly quantitative and that employ statistics and algorithms that many of their users may not completely understand…" (2015, p. 3). Marder argues that good intuitions about what methods of data analysis and manipulation ('data treatments') will provide "… an answer that is true to the essence of the biological process studied" (p. 3), are important for avoiding mistakes and making genuine progress. Her concerns are partly rooted in the problem of analytic flexibility and the sophistication of analysis techniques, but her primary concern is the growing body of evidence showing the various ways that brains are highly complex and interconnected systems. In particular, brains include multiple parallel pathways allowing for any given processing problem to be solved in multiple different ways (p. 2-3). Without good intuitions about what analysis technique to use when making sense of data about a system like this, it is likely that hypotheses will be pursued that, in fact, are not borne out in reality but only appear to be true as a consequence of complexity in the system under study. Marder argues that good models and theories that hone intuitions about "… which biological details are significant

for a given brain function and which details can, as a first approximation, be ignored" (p. 4) are necessary for effectively navigating this difficult situation. The view that familiarity with the material objects of investigation, in this case the biological materials that make up brains, is critical for effectively using data to study those objects is echoed in philosophical work on data sharing in biology (Leonelli 2013). Marder's concerns are also reinforced by recent discussions sparked by the replication crises occurring throughout the sciences.

The reuse potential of neuroimaging data, and the potential for progress provided by innovations in methods and approaches to analysis conspire to create an environment where inferences may not be verifiable. Neuroimaging research, because of the "…high dimensionality of fMRI data, the relatively low power of most fMRI studies and the great amount of flexibility in data analysis…" has recently been dubbed "… a 'perfect storm' of irreproducible results" (Poldrack et al 2017). The problem is framed as one of reproducibility, that is the ability to reconstruct the analysis and reasoning procedures used to arrive at a given body of evidence. Reproducibility has to do with the capacity of independent investigators to reconstruct a result from the original data following the same steps as reported by the original authors, and a common proposal for improving reproducibility is to foster transparent practices. Common suggestions include sharing data, sharing algorithms and analysis code, and the pre-registration of research plans, including methods of data analysis (Poldrack et al 2017; Munafò et al 2017). The success of these proposals depends on their adoption throughout the community, and their capacity to reveal the epistemically relevant aspects of the data interpretation process.

In summary, neuroimaging data are regarded as informationally rich, as they have the potential to be relevant for the study of phenomena beyond those they were produced to investigate. That potential is realized through the use and development of different data analysis techniques. The richness of neuroimaging data allows changes in the analysis process to influence the phenomena they are regarded as informative about. The richness of the data and variability in analysis processes is also a potential source of inferential errors. This potential and its realization through innovations in data analysis risks increasing false positive rates by increasing analytic flexibility, the depreciation of

intuitions based on details about the material objects under study, and reinforcing research practices that have made it difficult to independently reproduce experimental results. This is the potential and peril of data analysis in neuroimaging research as viewed from within. In the next section, I discuss the view of neuroimaging research from the perspective of philosophy.

1.3 The View from Philosophy

Philosophical work on neuroimaging research takes a variety of forms. This includes debates about the evidential value of fMRI data with respect to investigations of the relationship between the brain and cognition (Uttal 2001; Landreth and Richardson 2004; Aktunç 2014), the evidential value of the images produced by common analysis pipelines (Roskies 2010a; Klein 2010a), discussions about the validity (or invalidity) of particular inferential practices such as reverse inference (Machery 2014; Glymour and Hanson 2016), the ethical and social dimensions of fMRI research (e.g., Figdor 2013; Bluhm 2013), and debates about the adequacy of the current stock of concepts used to theorize about cognitive processes (Figdor 2011; Klein 2012; Anderson 2015). Here, I focus on the epistemic dimensions of fMRI research and the role data analysis plays in the data interpretation process. The original motivation for this project stems from an examination of skeptical arguments about neuroimaging research that persist in the philosophical literature, and in particular the recurring form of critical arguments and counter-arguments. The persistence and consistency of the back and forth between skeptics and critical advocates of neuroimaging technology leaves the impression that something important in the research practices is being overlooked. Situating the debate against the backdrop of the view from neuroscience presented in the previous section reinforces that impression.

Critics tend to examine an inference in neuroimaging research by outlining its logical structure, and then pick out one or more assumptions that must obtain for the inference to follow from neuroimaging data. These assumptions are then challenged, either by way of an argument from underdetermination that involves suggesting a number of alternative

possibilities or explanations for the results (e.g., Mole and Klein 2010), or by showing that the assumptions, once made explicit, reveal a vicious circularity (e.g., Hardcastle and Stewart 2002). Responses to skeptics also follow a similar pattern. They typically deflate the target skeptical argument by showing that the skeptics' presentation of the research is narrow (e.g., Landreth and Richardson 2004), misrepresents the data analysis techniques in questions (e.g., Machery 2014), or otherwise overlooks epistemically relevant aspects of the experimental practice (e.g., Roskies 2010b).

Criticisms of inferences common in neuroimaging research begin with van Orden and Paap's incisive critique of the logic of subtraction (1997). At the time this paper was published subtraction was the most used method of analysis in neuroimaging research. The method involves using two task conditions that differ by an isolable cognitive component. The example van Orden and Paap use is a task in which the subject examines pairs of words and decides whether or not they rhyme, and a contrast task in which the same subject examines pairs of words but does not indicate whether they rhyme. To perform a subtractive analysis, the BOLD signal data measured during each of these two tasks is subtracted. The resulting difference in BOLD activity is then attributed to the difference in cognitive activity. In the case they review, the subtraction isolated a small part of the left temporoparietal cortex as differentially more active during the rhyming judgement. They argue that this result does not warrant the conclusion that the cognitive process crucial for making rhyming judgements resides in the region picked out by the subtraction.

The use of subtraction, they argue, assumes that cognitive components can be isolated at all. This assumption, sometimes referred to as the 'pure insertion' hypothesis, reflects the notion that a single element or component of a cognitive process can be 'inserted' (or removed) from a process without effecting the overall performance of that process (Harrison and Pantelis 2010). Van Orden and Paap argue that "… the conclusion that an observed pattern of dissociated brain regions demonstrates separate cognitive components … simply affirms the inevitable consequent of assuming there were single causes in the first place…" (van Orden and Paap 1997, p. S90), and further that this assumption is likely to be false. This so-called 'doctrine of single causes', which

attributes a single cause to an observed effect, is "… at odds with the nature of cognitive systems" (p. S92), which they argue are reciprocally causal insofar as each component of a system contributes to every outcome of that system. Furthermore, they claim that, "[i]f the original assumption of single causes is false, the statistical tools will nevertheless discover components" (p. S90). Taken together, the use of subtraction to localize cognitive functions to parts of the brain is a failed program because the method assumes such localization is possible, will produce apparently successful localizations no matter the facts of the system, and forecloses viable alternative possibilities such as a distributed processing account of cognitive function. This argument, while the first, is not the only to challenge neuroimaging research in this way.

Valerie Hardcastle and Matthew Stewart provide a blanket criticism of research that claims to have localized any cognitive process to any discrete part of the brain that targets all of neuroscience, ranging from single cell recordings (2002, p. S73), to neuroimaging (p. S77). The argument runs the same course as van Orden and Paap's, resting on the observation that the methods of analysis, and research strategies more generally assume "… local and specific functions prior to gathering appropriate data for the claim" (p. S80). That is, the inferences from experimental results to claims about the cognitive contributions made by structures of the brain are based on experimental methods and analysis procedures that presuppose brain structures play a specific cognitive role. They reinforce their argument by an appeal to underdetermination, suggesting that the core problem is that results could be due to functional diversity of brain regions, details of the neurophysiology that are not considered when inferences are made, and cannot be considered given the assumptions required by the methods and techniques used to produce the data in the first place.

One common feature of these arguments is that they all take aim at subtraction analysis, which, as it turns out, is no longer the dominant method for analyzing and interpreting neuroimaging data. This change in analytic methods has, until recently, only been briefly alluded to in contributions to these debates (e.g., Klein 2010a; Roskies 2010a). A recent contribution does engage pattern classification analysis in detail, and it joins the skeptical chorus. Following the critical format outlined above, Brendan Ritchie, David Kaplan and

Colin Klein argue that interpreting pattern classification analysis results as evidence for claims about neural representations relies on assumptions about the cause of the classifier's performance that are not likely to be borne out (forthcoming).

These arguments share a similar structure. They claim that inferences in neuroimaging research are undermined by the assumptions motivating, or implicit in the use of, particular data analysis techniques. In each case, assumptions are shown to make the inferences viciously circular, or relevant and compelling possible alternatives are identified as overlooked by virtue of these assumptions. Compared against the perils of innovations in data analysis reviewed in the previous section, the novelty of these critiques is that they argue for a vicious circularity in the logic of the application of methods of data analysis. After all, the neuroscientific view on the challenges with interpreting neuroimaging data is attentive to the inferential risks that follow from the causal complexity of brain systems, and the indirect nature of neuroimaging data. If the inferential practices that skeptics identify accurately reflect the practice, then concerns about analytic flexibility and the inability to replicate experiments, are deeper problems than neuroscientists give them credit for. The other half of the debate about the epistemic status of neuroimaging data provides reason to doubt that skeptical arguments do in fact accurately capture the inferential practices neuroscientists engage in. Those responding to skeptical views tend to argue that the critics have, in one way or another, misrepresented or overlooked epistemically relevant dimensions of the experimental practice. I briefly review two such responses.

William Uttal's suggestively titled book The New Phrenology (2001) provides one of the most comprehensive critiques of the form noted above. In a more recent book that expands on the argument in the first, he argues that fMRI is an "epistemological sledge hammer" and that a close examination of the experimental protocols of a variety of publications is not encouraging: "[v]arious kinds of statistical manipulations may appear to define particular prototypical response patterns: however, given their variability all must be considered skeptically" (2011). In a response to Uttal's 2001 book, Anthony Landreth and Richard Richardson criticize him for misrepresenting neuroimaging methodology (p. 118). They note that, while Uttal is correct that separating signal from

noise in the BOLD signal is a difficult task, research practices such as controlling stimulus presentation, the repetition of trials to enable signal boosting through averaging, and the use of supplementary statistics, are essential aspects of the experimental process that Uttal doesn't take into account (p. 118-9). They conclude that, while the science is not perfect and Uttal raises important issues with respect to the prospects of localization projects, it isn't productive to raise those issues by misrepresenting the practice and leaving out epistemically relevant details.

Separately, Adina Roskies uses a similar strategy to refute the critique laid out by van Orden and Paap. She argues that interpretations of neuroimaging data involve what she calls 'functional triangulation', noting that "… in functional imaging, information from other task comparisons and other studies is brought to bear on the interpretation of experimental data" (Roskies 2010b, p 641), and further that "[c]onvergence across multiple experiments is key to epistemic warrant when it comes to attributing function to anatomical regions" (P. 641). The examples of convergence she notes includes convergence between different task paradigms within an experiment, across experiments within a discipline, and across different measurement techniques. Viewing the inferential practices engaged in neuroimaging research through the lens of functional triangulation, van Orden and Paap's criticism loses its force.[1] An individual experiment must be recognized as part of a larger practice. The inferential limitations of the results of subtraction analysis are not indicative of the limits of the entire data set, or domain of research. The consistent features of arguments resistant to skeptical conclusions, and the divergence between the pessimistic outlook of skeptics, and optimistic outlook of

---

[1] Hardcastle and Stewart's argument, which is critical of a number of different experimental methods common in neuroscience, can be diffused in a similar fashion. Their general skeptical view about neuroscience is based on arguments that each of the parts of neuroscience are inferentially limited. This is the very kind of argument, and conclusion, that Roskies' functional triangulation account of evidence resists. Additionally, with respect to neuroimaging in particular, they place the weakness of neuroimaging research on decisions about statistical thresholding in the application of subtraction (p. S78), and the fact that fMRI measures metabolic, not neural, change (p. S79). These are the very inferential challenges that, as I argue in what follows, neuroscientists engage through the use of a variety of data analysis techniques.

neuroscientists suggests that there may be a systematic problem in the way skeptics, and philosophers in particular, approach their analysis of neuroimaging data's evidential significance.

Neuroimaging research is not only discussed in the context of debates about the merits and challenges of neuroimaging data. Parallel to this debate is an argument by Jim Bogen that philosophical accounts of evidence in experiments do not capture what "make[s] experimental evidence" like that obtained from functional imaging technologies "epistemically valuable" (2002). Bogen focusses his attention on accounts of experimental evidence that emphasize error-statistics (Mayo 1996) and counter-factual dependencies between the processes of data production and claims data are used to support (Woodward 2000). Both of these views recognize the primary role of data manipulation as correcting for errors. Bogen observes that "... what is epistemically good about functional images is not that they are highly accurate with regards to [biological indicator] levels or locations of individual brains" (p. S65), and further that the purpose of manipulating imaging data is not to "... bring error-ridden, [biological indicator] estimates recognizably closer to what would have resulted from ideal experiments shielded from significant sources of error" (p. S65).

In the introduction of Bogen's paper he notes that neuroimages are different from the kinds of data considered when he and Woodward developed the data-phenomena distinction, which forms the basis of Woodward's counterfactual account of experimental evidence (2001, p.S61). The data they had in mind is data that owes their evidential value to "... having been produced in such a way that the item they are used to study exerts a detectable causal influence on them" (p. S61). Bogen suggests that neuroimages are distinct from data in that they are, as images, "… more like graphic representations of interpretations of data than what Woodward and I meant by data" (p. S61). Whether or not neuroimages themselves are best treated as data, or representations of interpretations, they are the most salient output of a neuroimaging experiment. It is common in the philosophical literature to treat the neuroimages produced as part of the analysis of neuroimaging data as the central piece of evidence produced by the analysis of neuroimaging data (e.g., Roskies 2010a; Klein 2010; Klein 2012; Machery 2014).

Whether this tendency involves treating neuroimages as a compact representation of the interpretive process, or as the final product of a complex data production process, it is important to recognize that neuroimages are just one of many data patterns used by investigators in the interpretation of neuroimaging data. This is the conclusion Klein arrives at in his aptly titled "Images Are Not the Evidence in Neuroimaging" (2010b).

Klein acknowledges that, because the brain is causally dense, that is a change in BOLD signal in one voxel can be caused by changes in almost any other part of the system, that neuroimages do not even provide weak evidence for claims about the cognitive contribution of regional brain activity (p. 275).[2] He cautions, however, against extending this to the whole of neuroimaging data. Neuroimages are the product of one analysis procedure, and they provide a 'first-pass sanity check' on the data, but additional analyses are required to interpret them (p. 275). Klein points towards 'more sophisticated analyses', details about neural anatomy and converging evidence for other research modes such as single cell recordings, as providing the additional evidential value to neuroimaging data above and beyond the neuroimages themselves (p. 276).

Evidence accumulated across laboratories and research paradigms is important for any account of explanation and knowledge production in neuroscience broadly construed (Bechtel 2004). Indeed, convergence across measurement technologies played a critical role in the eventual uptake of neuroimaging methods in cognitive neuroscience (Bechtel and Stufflebeam 1997). These factors certainly contribute to the strength of inferences in neuroimaging research, but they cannot explain what makes neuroimaging data itself epistemically valuable. They are part of the body of evidence neuroscientists appeal to

---

[2] It is noteworthy that this argument has been disputed for misrepresenting the kind of statistical inference neuroscientists engage in. Where Klein presents neuroimaging researchers as aiming to reject 'point null' hypotheses, statistical testing in neuroimaging research involves the rejection of a 'range null' hypothesis and not a point null, and the causal density of the system is not problematic in the way Klein makes it out to be for range null hypothesis tests (Machery 2014). That neuroscientists are concerned about the causal density of the brain, however, is reason to take Klein's argument seriously even if it misrepresents some of the technical aspects of the practice.

when making inferences, but neuroimaging data must also make their own contribution to the inferences if it is to be worth the effort to produce and analyze. Appealing to convergence with external results does not help clarify the intrinsic evidential value of neuroimaging data. This leaves 'sophisticated data analysis techniques' to make up the difference, which is the very thing that critics of the technology focus on in their analyses. That is, the same skeptical arguments that have been criticized for overlooking epistemically relevant aspects of the research practices neuroscientists engage in. The systematic problem with the skeptical approach, then, may be that they mistakenly treat all data analysis techniques as tools for reducing noise and 'approximating the results of an ideal experiment' (that is, an experiment that actually measured the causal factors of interest).

Some contributors to the discussion of evidence in neuroimaging have acknowledged that data analysis is important, and that techniques beyond subtraction (Roskies 2010a) and image production (Klein 2010a), make a difference to judgements of the evidential value of neuroimaging data. A detailed account of how data analysis techniques that are not used to account for error, eliminate artifacts or approximate the results of an ideal experiment, contribute to the evidential value attributed to neuroimaging data is at present absent. This leads to the primary question this project seeks to answer: What contribution do data analysis techniques like subtraction and pattern classification analysis make to the process of interpreting neuroimaging data? I pursue this question by addressing three narrower questions relating to the use of data analysis in neuroimaging research.

1.4 Towards a Philosophy of Data Analysis

The questions motivating each of the three papers that follow are: Is the skepticism about inferences in neuroimaging research warranted? How does the use of data analysis techniques account for the epistemic value neuroscientists recognize in neuroimaging data? How does the availability and use of data and analysis tools shape assessments of the evidential value of neuroimaging data? The included papers each examine how neuroscientists bring neuroimaging data to bear on claims about the relationship between cognition and the brain.

The first paper, “Data Analysis and the Evidential Scope of Neuroimaging Results”, argues that philosophical skepticism about the ability of neuroimaging data to support theoretical hypotheses that relate brain function to cognitive function is not warranted. I identify the skeptics’ strategy of treating data analysis techniques in isolation from the research context in which they are used as a shortcoming of the approach common to many skeptical arguments. To demonstrate the importance of situating the use of data analysis within a research context, I show how multiple data analysis techniques are used to minimize confounding interpretations of the analysis results. As it turns out, the assumptions addressed through the use of multiple analysis techniques include some of the assumptions that critics are keen to identify as sufficient for undermining these inferences.

The second paper, “The Interpretation of Neuroimaging Data as Explanations of Data Patterns”, presents a conceptual framework for evaluating the contribution data analysis techniques make to the interpretation of data. The problem with many skeptical arguments, as argued in the first paper and alluded to above, is that they are not attentive to the complexity of data interpretation, or the subtleties of the contribution data analysis techniques make to that process. By situating data analysis techniques within the process of data interpretation, I argue that data analysis techniques facilitate data interpretation by isolating data patterns that neuroscientist explain by appeal to claims about phenomena.

The third paper, “Data Analysis and The Perceived Value of Data", examines a dispute between neuroscientists over the evidential value of data represented in the NeuroSynth repository. NeuroSynth is novel insofar as it is curated and annotated by an algorithm. I use this dispute to argue that strategies for data interpretation are likely to be successful only insofar as they are sensitive to facts about how the data came to have the form that it does. In particular, I trace the inferential errors made by the users of the data to the inappropriate application of criteria of explanatory adequacy honed for the interpretation of locally produced data to the interpretation of synthesized data. Set alongside the first two papers, this paper begins exploring factors that contribute to the explanations of data patterns.

The concluding chapter takes up where this one leaves off: addressing the broad question set out above: how do data analysis techniques contribute to the inferential practices operative in neuroimaging research? To spoil the conclusion, I argue that they are not just tools for quantifying the strength of hypotheses via statistical tests, or correcting for various sources of noise and artifacts in the data. While some data analysis techniques perform these roles, the techniques at the heart of neuroimaging research make a distinct contribution to the data interpretation process. They are valuable because the results of data analysis techniques, unlike the 'raw' data provided by neuroimaging experiments, can be explained in terms of claims about phenomena. In this way data analysis techniques are integral to the practice of neuroimaging research. In the final chapter I use the arguments and cases examined in the included papers to add detail to this account, articulating more clearly the role of data analysis in the interpretive process in terms of the interpretive and epistemic leverage they provide.

References

Aktunç, E. M. [2014]: 'Severe Tests in Neuroimaging: What We Can Learn and How We Can Learn It', Philosophy of Science, 81, pp. 961-73.

Anderson, M. [2015]: 'Mining the Brain for a New Taxonomy of the Mind', Philosophy Compass, 10, pp. 68-77.

Bandettini, P. [2012]: 'Twenty years of functional MRI: The science and the stories', NeuroImage, 62, pp. 575-588.

Bechtel, W. P. [2004]: 'The epistemology of evidence in cognitive neuroscience', in R. Skipper Jr, C. Allen, R. A. Ankeny, C. F. Craver, L. Darden, G. Mikkelson & and R. Richardson (eds.), Philosophy and the Life Sciences: A Reader. MIT Press.

Bechtel, W. P. and Stufflebeam, R. S. [1997]: 'PET: Exploring the Myth and the Method', Philosophy of Science, 64, Supplement. Proceedings of the 1996 Biennial Meetings of the Phliosophy of Science Assosciation. Part II: Symposia Papers, pp. S95-S106.

Bluhm, R. [2013]: 'Self-Fulfilling Prophecies: The Influence of Gender Stereotypes on Functional Imaging Research on Emotion', Hypatia, 28, pp. 870-886.

Bogen, J. [2001]: 'Functional imaging evidence: Some epistemic hotspots' In Peter K. Machamer, Peter McLaughlin & Rick Grush (eds.), Theory and Method in the Neurosciences. University of Pittsburgh Press. pp. 173-199.

Bogen, J. [2002]: 'Epistemological custard pies from functional brain imaging', Philosophy of Science, 69, pp. S59-S71.

Bucker, R. L., Andrews-Hanna, J. R., and Schacter, D. L. [2008]: 'The Brain's Default Network: Anatomy, Function, and Relevance to Disease', Ann. N. Y. Acad. Sci., 1124, pp. 1-38.

Carp, J. [2012]: 'On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments', Frontiers in Neuroscience, 6, 149.

Coutanche, M. N. [2013]: 'Distinguishing multi-voxel patterns and mean activation: Why, how, and what does it tell us?', Cognitive Affective Behavioural Neuroscience, 13, 667-673.

Figdor, C. [2013]: 'What is the "Cognitive" in Cognitive Neuroscience?', Neuroethics, 6 (1), pp.105-114

Figdor, C. [2011]: 'Semantics and Metaphysics in Informatics: Towards an Ontology of Tasks', Topics in Cognitive Sciences, 3, pp. 222-226.

Gelman, A. and Loken, E. [2013]: 'The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time∗', Department of Statistics, Columbia University.

Glymour, C. and Hanson, C. [2016]: 'Reverse Inference in Neuropsychology', British Journal for the Philosophy of Science, 67, pp. 1139-1153.

Gusnard, D. A., and Raichle, M. E. [2001]: 'Searching for a baseline: Functional imaging and the resting human brain', Nature Reviews Neuroscience, 2, pp. 685-694.

Hardcastle, V. G. and Stewart, M. C. [2002]: 'What do brain data really show?', Philosophy of Science, 69, pp. S72-82.

Harrison, B. J, and Pantelis, C. [2010]: 'Cognitive subtraction', Encyclopedia of Psychopharmacology, pp. 323.

Haxby, J. V. [2010]: 'Multivariate Pattern Analysis of fMRI data' in M. Bunzl and S. J. Hanson (eds), Foundational Issues in Human Brain Mapping, The MIT Press, pp. 55-68.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai A., Scouten, J. L., and Pietrini, P. [2001]: 'Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex', Science, 293, pp. 2425-30.

Ioannidis, J. P. A. [2005]: 'Why most published research findings are false", PLoS Medicine, 2(8), e124.

Kanwisher, N. [2017]: 'The Quest for the FFA and Where It Led', The Journal of Neuroscience, 37, pp. 1056-1061.

Kamitani, Y., and Tong, F., [2005]: 'Decoding the visual and subjective contents of the human brain', Nature Neuroscience, 8, pp. 679-685.

Klein, C. [2012]: 'Cognitive Ontology and Region- versus Network-Oriented Analyses', Philosophy of Science, 79(5), pp. 952-960.

Klein, C. [2010b]: ‘Philosophical issues in neuroimaging’, Philosophy Compass, 5, pp. 186-98.

Klein, C. [2010b]: ‘Images are not the evidence in Neuroimaging’, British Journal for the Philosophy of Science, 61, pp. 265-78.

Kohler, P.J., Fogelson, S.V., Reavis, E.A., Meng, M., Guntupalli, J.S., Hanke, M., Halchenko, Y.O., Connolly, A.C., Haxby, J.V. and Peter, U.T., [2013]: ‘Pattern classification precedes region-average hemodynamic response in early visual cortex’. NeuroImage, 78, pp. 249-260.

Kriegeskorte, N. and Kievit, R. A. [2013]: ‘Representational geometry: integrating cognition, computation, and the brain’, Trends in Cognitive Sciences, 17, pp. 401-12.

Landreth, A. and Richardson, R. C. [2004]: ‘Localization and the new phrenology: A review essay on William Uttal’s the new phrenology’, Philosophical Psychology, 17, pp. 107-23.

Leonelli, S. [2013]: ‘Data Interpretation in the Digital Age’, Perspectives on Science, 22, pp. 397-417.

Liu, T., Hospadaruk, L., Zhu, D. C., and Gardner, J. L. [2011]: ‘Feature-Specific Attentional Priority Signals in Human Cortex’, The Journal of Neuroscience, 31, pp. 4484-95.

Machery, E. [2014]: ‘Significance Testing in Neuroimagery’, in Kallestrup J., and Sprevak, M.(eds.), New Waves in the Philosophy of Mind, Palgrave Macmillan, pp. 262-277.

Marder, E. [2015]: ‘Understanding Brains: Details, Intuitions, and Big Data’, PLoS Biology, 13(5), e1002147.

Martin, C. B., Cowell, R. A., Gribble, P. L., Wright J., and Köhler S. [2015] ‘Distributed category-specific recognition memory signals in human perirhinal cortex.’ Hippocampus.

Martin, C. B., McLean, D. A., O’Neil, E. B., and Köhler S. [2013]: ‘Distinct Familiarity-Based Response Patterns for Faces and Buildings in Perirhinal and Parahippocampal Cortex’, The Journal of Neuroscience, 33, pp. 10915-23.

Mayo, D. [1996]: Error and the Growth of Experimental Knowledge, The University of Chicago Press.

Mole, C. and Klein, C. [2010]: ‘Confirmation, Refutation, and the Evidence of fMRI’, In Stephen Hanson & Martin Bunzl (eds.), Foundational Issues in Human Brain Mapping. Cambridge: MIT Press. pp. 99-112.

Morcom, A. M., and Fletcher, P. C. [2007]: ‘Cognitive neuroscience: The case for design rather than default’, Neuroimage, 37, pp. 1097-1099.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E., Ware, J. J., and Ioannidis, J. P. A. [2017]: ‘A manifesto for reproducible science’, Nature Human Behaviour, 1.

Norman, K., Polyn, S. M., Detre, G. J., and Haxby, J. V. [2006]: 'Beyond mind-reading: multi-voxel pattern analysis of fMRI data', TRENDS in Cognitive Sciences, 10(9), pp. 424-430.

Pereira, F., Mitchell, T., and Botvinick M. [2009]: 'Machine learning classifiers and fMRI: A tutorial overview', Neuroimage, 45, p. 199-209.

Poldrack, R. A. [2011]: 'Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding', Neuron, 72(5), pp. 692-697.

Poldrack, R. A. [2006]: 'Can cognitive processes be inferred from neuroimaging data?', Trends in Cognitive Sciences, 10(2), pp. 59-63.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò M. R., Nichols, T. E., Poline, J., Vul, E., and Yarkoni, T. [2017]: 'Scanning the horizon: towards transparent and reproducible neuroimaging research', Nature Reviews Neuroscience, 18, pp. 115-126.

Reddy L., Tsuchiya N., and Serre T. [2010]: 'Reading the mind's eye: decoding category information during mental imagery', Neuroimage, 50(2), pp. 818-825.

Ritchie, J.B., Kaplan, D.M., and Klein, C. [Forthcoming]: 'Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience', British Journal for the Philosophy of Science.

Roskies, A. [2010a]: 'Neuroimaging and Inferential Distance: The Perils of Pictures', in M. Bunzl and S. J. Hanson (eds), Foundational Issues in Human Brain Mapping, The MIT Press, pp. 195-216.

Roskies, A. [2010b]: 'Saving Subtraction: A reply to Van Orden and Paap', British Journal for the Philosophy of Science, 61, pp. 635-65.

Sandberg, K., Andersen, L. M., and Overgaard, M. [2014]: 'Using multivariate decoding to go beyond contrastive analysis in consciousness research', Frontiers in Psychology, 5, pp. 8-13.

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud G., Duff, E., Feinberg, D. A., Giffanti, L., Harms, M. P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A. Z., Vu, A. T., Woolrich, M. W., Xu, J., Yacoub, E., Uğurbil, K., Van Essen, D. C., Glasser, M. F., for the WU-Minn HCP Consortium. [2013] 'Resting-state fMRI in the Human Connectome Project', NeuroImage, 80, pp. 144-168.

Tambini, A., and Davachi, L. [2013]: 'Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory', Proceedings of the National Academy of Sciences of the United States of America, 110(48), pp. 19591-19596.

Tong, F., and Pratte, M. S. [2012]: 'Decoding Patterns of Human Brain Activity', Annual Review of Psychology, 63, pp. 483-509.

Uttal, W. [2001]: The New Phrenology, The MIT Press.

Van Horn, J. D., and Gazzaniga M. S. [2013]: 'Why share data? Lessons learned from fMRIDC', Neuroimage, 82, pp. 677-682.

van Orden, G. C., and Paap, K. R. [1997]: ‘Functional Neuroimages Fail to Discover Pieces of Mind in Parts of the Brain’, Philosophy of Science, 64, pp. S85-94.

Walther, D. B. [2013]: ‘Using confusion matrices to estimate mutual information between two categorical measurements’, Proceedings of the 3rd International Workshop on Pattern Recognition in NeuroImaging, p. 220-224. Philadelphia, PA.

Woodward J. [2000]: ‘Data, phenomena, and reliability’, Philosophy of Science, 67(3), pp. S163-S179.

Chapter 2

2 The Analysis of Data and the Evidential Scope of Neuroimaging Results

2.1 Introduction

The debate amongst philosophers about the epistemic status of neuroimaging begins with van Orden and Paap's criticism of the logic of subtraction (1997), the primary technique used to analyse neuroimaging data at the time their paper was published. Philosophers have continued to debate the strengths and weaknesses of neuroimaging as a tool for investigating the relationship between cognitive functions and the brain (Uttal 2001; Hardcastle and Stewart 2002; Roskies 2010a; Klein 2010a; Uttal 2011; Aktunç 2014). I argue that, since most critics have not taken into account the significance of the diversity of data analysis techniques used to analyse neuroimaging data, the skepticism towards neuroimaging technology is misplaced.

Many of the skeptical positions are grounded on careful analyses of subtraction and subtraction logic (Uttal 2001; Hardcastle and Stewart 2002; Klein 2010a). While philosophers are rightly critical of the ability of subtraction analyses, on their own, to support claims about the relationship between cognitive functions and the brain, subtraction is only one kind of data analysis technique used to analyse neuroimaging data. Given that the development of new data analysis techniques has been a significant driver of progress in neuroimaging over the last decade and a half,[3] a narrow focus on

[3] There has been a steady shift from using univariate analysis techniques that treat the neuroimaging data as a scalar value, usually an average, towards the use of multivariate analysis techniques that treat the neuroimaging data as a vector. These new techniques have allowed neuroimaging researchers to pursue new theoretical goals and study new

subtraction is a problem for any argument that aims to shed light on the range of hypotheses that neuroimaging technology can discriminate between.

Indeed, some recent contributors have noted that the role and impact of multivariate analyses has not been fully appreciated in this debate (Roskies 2010a; Klein 2010b). However, while they acknowledge that techniques other than subtraction are important to consider, they do not, themselves, take up the task of exploring how the use of other analysis techniques changes the evidence available in neuroimaging research. My aim here is to begin to fill this gap by demonstrating that, when evaluating the hypotheses and claims that neuroimaging technology can and cannot support, it is important to take into account the contribution of new analysis techniques such as pattern classification analysis, and to consider how multiple analysis techniques can be brought together to strengthen the evidence provided by neuroimaging technologies.

I proceed as follows: In section two I review the debate about the epistemic status of neuroimaging and specify the categories of hypotheses that philosophers claim neuroimaging data can and cannot support. In section three I present a conceptual framework for evaluating the strength and content of evidence produced via a data analysis technique. In section four I apply this conceptual framework to a study that uses multiple analysis techniques to generate evidence in support of a hypothesis and, where critics of neuroimaging would argue that the data do not support this hypothesis, I show how they can. The evidence is stronger than it appears because one analysis technique is used to validate a crucial assumption required by the other. In section five I argue that different analysis techniques provide different evidence, and that the use of multiple analysis techniques to examine the same data provides experimental results with a kind of local robustness.

---

hypotheses, such as the investigation of the content of neural representations (see Tong & Pratte 2012 for an introductory review of multivariate techniques).

2.2 Skepticism About Neuroimaging

Functional magnetic resonance imaging (fMRI) allows neuroscientists to study the human brain through non-invasive measurements of metabolic activity (see Ashby 2011 for a technical introduction). Experiments using fMRI typically require a participant to perform a cognitive task – such as identifying faces as familiar or unfamiliar (as in Martin et al. 2013) – while the scanner measures changes in the Blood Oxygenation Level Dependent (BOLD) signal throughout their brain. [4] Roughly speaking, the scanner does this by dividing the brain into voxels (volumetric pixels), which are one millimeter to three millimeter cubes of brain matter, and measuring the BOLD signal in each voxel over time. The value of the BOLD signal is the ratio of oxygenated to deoxygenated hemoglobin in a voxel at the time of scanning. Since it tracks properties of blood flow, the BOLD signal is often referred to as the hemodynamic signal. After a scanning session, the investigators will have a data set that consists of BOLD signal values for each voxel labeled with the task condition that the participant was performing when that data was collected.

Neuroimaging data has historically been analysed using subtractive analyses. In the simplest case of subtraction, two sets of neuroimaging data are required, each obtained while the participant performs a different task. The goal of subtractive analysis is to identify the difference in BOLD signal that corresponds with the cognitive difference between the tasks. Roughly, the BOLD signal values in each voxel associated with one task are subtracted from the values in the same voxels associated with the other task. This analysis is classified as univariate because each voxel is treated independently of each

---

[4] The fMRI scanning protocol does not directly measure metabolic activity. During an fMRI scan, radio pulses cause hydrogen atoms to align with a uniform magnetic field. As they relax to equilibrium they release energy, which the scanner measures. Deoxygenated hemoglobin, unlike oxygenated hemoglobin, causes the nearby magnetic field strength to vary, resulting in a difference in the measured energy, and forming the basis of the BOLD signal.

other voxel, and as such the data needs to be corrected for multiple comparisons. The result is a difference map that identifies the voxels (or regions) where brain activity is significantly different between the task conditions. To evaluate if the difference effects can be attributed to the population, and not just a given subject in the study, a second-level analysis is carried out (typically random effects analysis, see Friston et al 1999). If this analysis shows the difference to be consistent across subjects, then the cognitive difference between the tasks is attributed to the regions of the brain shown to be differentially active. To illustrate the conceptual logic of the process, consider van Orden and Paap's toy example in which task A is reading two words, and task B is reading two words then judging whether or not they rhyme (1997). The resulting subtraction between the imaging data obtained during task A and the data obtained during task B is taken to indicate the regions of the brain that are involved in the cognitive process that underlies the rhyming judgment.

Van Orden and Paap argue that the subtractive method cannot be used to locate where in the brain cognitive functions 'reside' because the reliability of subtractive inferences depends on several assumptions that they believe are not likely to be true. In particular, the reliability of subtraction with respect to localizing cognitive functions to regions of the brain requires that '. . . one must begin with a "true" theory of cognition's components, and assume that the corresponding functional and anatomical modules exist in the brain' (1997, p. S86). These assumptions, they argue, follow from the fact that a valid subtraction requires that the task-difference precisely isolates a single cognitive component, which can only be the case if the cognitive theory used to design the tasks is accurate (p. S87). Additionally, they argue that functional localization using subtraction further requires that those modules are feed-forward '. . . to insure that the component of interest makes no qualitative changes "upstream" on shared components of experimental and control tasks' (p. S86), and that the contrasted tasks '. . . invoke the minimum set of components for successful task performance' (p. S86).

William Uttal engages in a similar kind of skeptical attack on neuroimaging in his book (2001). Building on van Orden and Paap's critique, Uttal compares neuroimaging to phrenology and argues, among other things, that it requires the false assumption that

cognitive processes are managed and maintained by isolable modules of the brain. Valerie Hardcastle and Matthew Stewart (2002), express a similar type of skepticism, arguing that the logic of neuroimaging is viciously circular and conclude that '. . . neuroscientists cannot use the data they get to support their claims of function . . .' because '. . . they are assuming local and specific functions prior to gathering appropriate data for the claim' (p. S80). These critiques all point to a vicious circularity in the inference from the results of subtraction analysis to claims about the localization of cognitive function.

Some philosophers have defended cognitive neuroscience from these criticisms. For instance, Landreth and Richardson responded to Uttal's arguments (2004) in part by clarifying the details of how neuroimaging data are processed, analysed and interpreted. Additionally, Roskies has rejected van Orden and Paap's characterization of subtraction (2010b). She argues that subtraction results are just one part of a more complex scientific procedure that she calls functional triangulation, whereby '. . . information from other task comparisons and other studies is brought to bear on the interpretation of experimental data' (p. 641). She also argues that characterizing neuroimaging as solely aimed at localizing cognitive functions to specific brain regions, as the three critics noted above do, is not representative of all uses of neuroimaging data. After providing examples of the variety of theoretical aims neuroimaging and subtraction methods are put towards she concludes that 'without recognizing the diversity of the immediate goals of imaging studies, it is impossible to do justice to the technique' (p. 639).

Indeed, the recent development of new multivariate analysis techniques,[5] which were introduced to discriminate between modular and distributed accounts of the role that the

---

[5] It is important to note that the techniques discussed here, collectively referred to as multivariate pattern analyses (MVPA), are neither the only nor first multivariate techniques to be used in neuroimaging. For example, spatio-temporal partial least squares (PLS) is a multivariate technique that has been in use since the late 90s (McIntosh *et al.* 1996; 1998). I owe this clarification to an anonymous reviewer.

ventral visual pathway plays in visual perception (Haxby et al. 2001), has motivated cognitive neuroscientists to investigate hypotheses about the content of brain activity. In a review of the theoretical uses of multivariate techniques the authors predict that '… the enhanced sensitivity and information content provided by these methods should greatly facilitate the investigation of mind-brain relationships by revealing both local and distributed representations of mental content, functional interactions between brain areas, and the underlying relationships between brain activity and cognitive performance' (Tong and Pratte 2012, p. 503). The study of mental content, neural representations and the characterization of these in terms of distributed patterns of brain activity are very different theoretical goals than the localization of cognitive functions to parts of the brain. This is grist for Roskies' mill. Whenever critics of neuroimaging research treat it solely in terms of localization, the critics have failed to appreciate the variety of theoretical applications that the technology is put towards. Furthermore, this theoretical shift, which was made possible by the development of data analysis techniques that treat neuroimaging data as multidimensional patterns, illustrates the importance of evaluating analysis techniques other than subtraction when evaluating the epistemic value of neuroimaging technology.

Despite these defenses of neuroimaging, and the theoretical and analytic advances in the field of cognitive neuroscience, the general trend towards skepticism and the focus on subtractive analyses has persisted. While more recent conclusions tend to be on the milder side of skepticism, philosophers continue to challenge the ability of neuroimaging technology to provide evidence that supports the claims neuroscientists use the technology to investigate. Additionally, they continue to do so on the basis of an evaluation of subtraction and subtraction logic. I will examine one of the most recent contributions to this debate in more detail, as it challenges the inferences neuroscientists make on the basis of an evaluation of subtraction and subtraction logic, and leans on the rest of the skeptical literature to reinforce its conclusions (Aktunç 2014). In line with the skeptical tradition, Aktunç argues that, while neuroimaging data are useful, they cannot be used to support the kinds of hypotheses that cognitive neuroscientists use them to support.

Aktunç distinguishes between two types of hypotheses that neuroimaging data might be brought to bear on. There are hemodynamic hypotheses, which relate BOLD signal activity to the performance of cognitive tasks, or parameters of the tasks. There are also theoretical hypotheses, which relate cognitive processes to the brain structures that implement them (this distinction is from Huettel et al. 2008). To illustrate this distinction consider the following example: The claim that patterns of BOLD signal activity in both PrC and PhC are sensitive to differences between faces, buildings and chairs (Martin et al. 2013, p. 10921), is a hemodynamic hypothesis. The tasks used in this study require participants to judge images of faces, buildings and chairs as familiar or novel. Thus, this claim is about the relationship between patterns of BOLD signal activity and features of stimuli used in the cognitive task that participants performed. After discussing these results, the researchers advance a theoretical hypothesis. They claim that the '. . . findings indicate that both PrC and PhC contribute to the assessment of item familiarity' (p. 10922). This is a theoretical hypothesis because it identifies two brain structures, PrC and PhC, and specifies a cognitive process that they implement, the assessment of item familiarity. It is worth noticing the inferential relationship between these two types of hypotheses: the theoretical hypothesis is inferred from the hemodynamic hypothesis. Where a hemodynamic hypothesis specifies BOLD signal activity, a theoretical hypothesis specifies a structure of the brain. Likewise, where a hemodynamic hypothesis specifies a cognitive task, a theoretical hypothesis specifies a cognitive process.

Given this distinction between hemodynamic and theoretical hypotheses, Aktunç uses Deborah Mayo's error statistical framework to argue that neuroimaging data can only provide a severe test of hemodynamic hypotheses. On the simplest interpretation of Mayo's severity criterion, a hypothesis passes a severe test just in case (1) the data agree with the hypothesis and (2) there is a sufficiently high probability that, if the hypothesis were false, then the data would not agree with the hypothesis (Mayo 2005, p. 99).

Aktunç argues that, while neuroscientists may be interested in providing evidence that supports theoretical hypotheses, neuroimaging only has evidential import with respect to hemodynamic hypotheses (2014, p. 969). This is because a difference in mean BOLD signal, which is the pattern identified by subtractive analyses, can be embedded in a

statistical significance test. From this, Aktunç argues that '. . . using error probabilities, we can find out whether specific fMRI experiments constitute a severe test of specific hemodynamic hypotheses. Thus, fMRI data do have evidential import for hemodynamic hypotheses' (p. 969). His argument that theoretical hypotheses cannot be subjected to severe testing relies on two premises. First, there is the 'fact' that '. . . fMRI obviously does not test for the existence of cognitive modules or functions as defined by theories of cognitive science' (p. 969) because '. . . fMRI gives us data only on hemodynamic activity . . .' (Aktunç 2014, p. 968). The second premise rests on the arguments made in the existing skeptical literature (specifically Uttal 2001; Hardcastle and Stewart 2002; Klein 2010a). Thus, according to Aktunç, neuroimaging data cannot support theoretical hypotheses because (1) the data are indirectly related to the content of those hypotheses and, (2) critiques of subtraction analysis show that such inferences are viciously circular, unstable or otherwise unreliable. Neither of these premises can support the derived conclusion.

Inferences from neuroimaging results to theoretical hypotheses, like most inferences from measurement results to theoretical claims, are ampliative; hemodynamic activity is at best an indirect measure of neural activity (see Logothetis 2008), and task performance is at best an indirect indicator of cognitive functions (see Poldrack 2010a). However, the indirect relationship between the data and content of the theoretical hypothesis is not sufficient to support the claim that neuroimaging cannot provide evidence for hypotheses that relate cognitive functions to brain activity. Whether or not these inferences are warranted depends on the particular theoretical hypotheses that are advanced, and whether or not the assumptions required by the inferences are justified. Indeed, this is how van Orden and Paap originally argued against the logic of subtraction. It was not on the basis of the indirectness of the data itself, but on the basis of the specific assumptions required to infer from the data to a theoretical hypothesis of a certain kind.

However, no matter where you stand on the reliability of inferences from subtraction analysis to claims about the localization of cognitive functions, these arguments cannot be grounds for a sweeping claim about the evidential scope of neuroimaging data. Just because one data analysis technique has certain limitations does not mean that the data

themselves are similarly limited. Indeed, neuroimaging data can be, and are, analysed with other analysis techniques that reveal different patterns and correlations in the data. Whether or not neuroimaging data provides evidence in support of theoretical hypotheses depends on how the other analysis techniques help neuroscientists to mediate the inferential gap between hemodynamic hypotheses and theoretical hypotheses.

Inferences to theoretical hypotheses from neuroimaging data can be, and in practice are, strengthened by the use of multiple analysis techniques. The specific case I consider is when analysis techniques are used in sequence as a way to validate assumptions required by the primary analysis procedure. In the final section I distinguish this use of multiple analyses from functional triangulation as discussed by Roskies, in which multiple independent analyses provide convergent evidence for a hypothesis. In the next section I provide a framework for evaluating the kinds of information about theoretical hypotheses data analysis techniques provide.

2.3 Data Analysis and Evidence

The skeptical position reviewed in the previous section is a claim about the kinds of hypotheses neuroimaging data can and cannot support. According to skeptics, it can support hemodynamic hypotheses, which specify a relation between features of the data. It cannot support theoretical hypotheses, which specify a relation between the phenomena that those features are taken to indicate. Whether it is used to investigate a hemodynamic or theoretical hypothesis, neuroimaging data needs to be manipulated to reveal relationships between features of the data that are relevant to the hypothesis under investigation. This is the function of data analysis techniques, such as subtraction and pattern classification analysis.

Data analysis techniques transform the data produced by experimentation into evidence suitable for statistical analysis. These transformations reveal patterns and correlations between features of the data, which are then taken to be evidence in support of a hypothesis. Bogen and Woodward's distinction between data and phenomena (1988) is a useful place to begin thinking about this process. Broadly speaking, they characterize data, which are the result of the interaction between experimental design, implementation

and measurement, as '. . . idiosyncratic to particular experimental contexts, and typically cannot occur outside of those contexts' (p. 317). Phenomena, on the other hand '. . . have stable, repeatable characteristics which will be detectable by means of a variety of different procedures, which may yield quite different kinds of data' (p. 317). On this view, data provide evidence for claims about phenomena, while claims about phenomena provide evidence for theories.

Bogen and Woodward illustrate this by considering how one might determine the melting point of lead (pp. 309-310). To do so, a researcher might take several measurements of a sample of lead just after it melts. The data, in this case, is a collection of temperature measurements. These temperature measurements provide evidence about the melting point of lead, which is a claim about a phenomenon. The data are idiosyncratic because the result of each temperature measurement depends on a complex network of causal interactions, many of which are not related to the phenomenon of interest. The value of each temperature measurement will be influenced by features of the thermometer used, the heating apparatus, the sample of lead, the time of day, the ambient temperature, and more additional causal factors than could be named. After collecting sufficiently many measurements, the researcher averages them and, on the basis of the value of that average, makes a claim about the melting point of lead. Notice that it is not the individual temperature measurements, but the average value of the temperature measurements that provides evidence in support of a claim about the melting point of lead. This calls attention to a general feature of scientific practice: the individual data points, which are the products of specific runs of an experiment, need to be transformed to reveal their evidential value. Typically, this involves eliminating the effects of factors that contribute to the value of specific data points that are not relevant to the theoretical question or hypothesis under investigation. The data, without the influence of these factors removed, speaks only to the melting point of this sample of lead, at this time, as measured with this thermometer. Factors such as those arising from the peculiar features of the thermometer used are irrelevant to the melting point of lead insofar as they distort or conceal patterns in the data that reflect the 'true' melting point of lead.

After data is produced it is manipulated so that the patterns relevant to the phenomenon of interest are revealed and the irrelevant patterns are suppressed. Averaging the temperature measurements of melted lead is intended to suppress the patterns in the data caused by the irrelevant causal factors that contribute to the value of each specific data point. Other examples of manipulations that suppress irrelevant patterns are noise reduction procedures, and manipulations that remove the effect of measurement artifacts. Averaging, as well as more complex analytic techniques such as those discussed in detail below, transform data such that patterns relevant to the phenomenon in question are revealed. The result of these manipulations is taken to be evidence for one or more claims about the phenomenon. A data analysis technique, then, is a series of data manipulations, or transformations, that clarify the evidential import of the data. [6]

Different data analysis techniques can be distinguished by the data points that they operate on and by the specific transformations of the data they involve. For example, univariate and multivariate techniques can be distinguished by the data points that they manipulate. Univariate techniques, such as subtraction, treat voxels as independent variables while multivariate techniques, like pattern classification analysis (discussed in detail below) and representational similarity analysis (Kriegeskorte and Kievit 2013), treat the data as having many dependent variables. Data analysis techniques that operate on the same class of data points, such as these two multivariate techniques, can be distinguished by the particular manipulations they apply to the data. For example, pattern classification analysis uses a machine-learning decision procedure to classify the data, whereas representational similarity analysis uses a measure of similarity to compare brain activity between task conditions.

Data manipulations are important because they transform otherwise complex data into a form that investigators can interpret and statistically analyse[7] (Good 1983, pp. 285-286).

---

[6] Thanks to an anonymous reviewer for this phrasing.

[7] This process is often referred to as data reduction.

Each manipulation, by virtue of the transformation that it makes, imposes assumptions on the result. These assumptions limit what the result can be taken as evidence about. Just as van Orden and Paap identified several assumptions required by the use of subtractive analyses, most data manipulations require researchers to make assumptions about the data. For example, a standard manipulation performed on neuroimaging data is the removal of patterns caused by magnetic field drift. Magnetic resonance scanners use the variations in a magnetic field to detect the BOLD signal, and the magnetic field in some scanners slowly changes during the course of scanning. Manipulating data such that the effects of field drift are removed requires assuming that the data are corrupted by magnetic field drift. If the procedure is used on data produced by a scanner that does not have a field drift, then the procedure would introduce artificial patterns into the data. It would create artificial patterns in the data because the required assumption, that the scanner has a field drift with specific parameters, is not true of the data. In the case of field drift correction, the assumption can be validated by measuring the field drift of a scanner. This simple example illustrates how data manipulations entail or require assumptions to be made of the data, and shows that treating a specific data manipulation in isolation from the rest of the experimental process can make the evidential status of the data appear weaker than it in fact is.

Different analysis techniques operate on different data points, implement different manipulations and require making different assumptions of the data. This is how they reveal (and suppress) different data patterns. For example, subtraction reveals correlations between average amplitudes of the BOLD signal and task performance. Techniques like subtraction, when they include processes for smoothing and averaging the signal, suppress information about differences in activity between voxels within a region. Thus, some subtraction analyses are unable to reveal correlations between the co-ordinated activity of groups of voxels that preserve the same level of average activity between tasks. On the other hand, multivariate techniques, such as pattern classification analysis, correlate distributed patterns of BOLD signal activity with task performance. Pattern classification analysis is sensitive to distributed activity patterns that univariate techniques, like subtraction, cannot detect. However, multivariate techniques are less sensitive to one-dimensional effects that covary with stimulus features, to which

univariate techniques are very sensitive (see Davis and Poldrack 2013 for a detailed discussion of the uses of these techniques).

By leveraging their differences, investigators can use several data analysis techniques together to overcome the inferential limitations of a particular technique. The limitations of a technique tend to derive from the assumptions that the technique requires. If assumptions can be identified, depending on the nature of those assumptions, other data analysis techniques can be used to validate them. In this way, the use of multiple analysis techniques on the same data can strengthen an inference from the result of one analysis to the target hypothesis by providing a clearer picture of the evidential import of the data.

Specifically, where a given analysis technique provides evidence that can support a hemodynamic hypothesis, the inference from that hypothesis to a theoretical hypothesis will require investigators to make further assumptions about the data. Since different data analysis techniques reveal different patterns, it is often possible to validate some of those assumptions by analyzing the data in another way. This is how multiple analysis techniques can come together to strengthen the inference from a hemodynamic to a theoretical hypothesis. Typically, this is done through functional triangulation (Roskies 2010a), where multiple techniques are used separately on the data, and the hypotheses inferred are further supported by independent analysis of different data sets. The case I will discuss below is different, as the evidence is strengthened not through the independent application of multiple analyses, but the sequential application of analysis techniques.

2.4 Case: Deconvolution and Pattern Classification Analysis

Liu and colleagues' study aims to determine the role that certain regions of the brain play in directing attention (Liu et al. 2011). The primary analysis technique used is pattern classification analysis, a multivariate technique derived from research on machine learning. Pattern classification analysis is used to determine if cognitive tasks can be differentiated exclusively on the basis of patterns in the BOLD signal that correlate with task performance. As I argue below, this technique alone cannot support a theoretical

hypothesis attributing a cognitive role to activity within a region or part of the brain. However, Liu and colleagues do not deploy the technique in isolation. Their analysis includes a region of interest selection procedure that partially validates one of the crucial assumptions required by pattern classification analysis. While this does not provide definitive evidence in support of the theoretical hypothesis they advance, it demonstrates how multiple techniques can be used together to bring neuroimaging data to bear on hypotheses beyond those that merely relate hemodynamic activity to task performance.

Two behavioural tasks were used to generate their data set. In both tasks subjects were presented with two overlaid patterns of dots and were instructed to attend to one pattern or the other. In the first task both patterns were composed of white dots, but one was rotating clockwise and the other counterclockwise. In the second task, both patterns were moving in a random-walk, but one was composed of red dots and the other green dots (p. 4485-6). The resulting data set contained BOLD signal measurements for each of the six task conditions: attending to clockwise rotating dots, attending to counter-clockwise rotating dots, attending to red dots, attending to green dots and the null-condition for each task (attending to a fixation cross).

The data were pre-processed before they were analysed. This involved head motion correction (to remove artifacts caused by subjects moving while being scanned), removal of low-frequency drift (this corrects for a scanning artifact due to a drift in the magnetic field of the scanner) and conversion of the BOLD signal measurements from raw values into a percentage of signal change (p. 4486). The result of these transformations is a data set suitable for the analysis procedures with patterns due to known artifacts from head motion and scanner drift suppressed. The pre-processed data were analysed using a series of analysis techniques. Before discussing the techniques in detail, I will provide a brief overview of the whole procedure.

The analysis began with deconvolution, a technique used to isolate the task-relevant portion of the BOLD signal data. The result of the deconvolution analysis was used as the input for a region of interest (ROI) selection procedure. The combination of the deconvolution and ROI selection was then used as the input for pattern classification

analysis. The result of the pattern classification analysis was then taken to support a claim about the regions of the brain involved in the modulation of attentional control. Notice that this is not a claim about the relationship between task performance and hemodynamic activity. It is a claim about which parts of the brain implement a particular cognitive process (modulation of attentional control). It is about the relationship between a cognitive function and regional brain activity. This is a theoretical hypothesis.

There are multiple inferences involved in moving from a hemodynamic hypothesis to a theoretical hypothesis. Recall that a hemodynamic hypothesis relates BOLD signal data to the performance of a task, whereas a theoretical hypothesis relates brain structure (or the activity in brain structure) to a cognitive process. Inferring from one to the other requires treating the BOLD signal measurements as an indicator of cognitively relevant brain activity within a brain structure, and task performance as an indicator of one or more cognitive processes. Whether or not the task can be taken as an indicator of the cognitive function that the researchers are interested in depends on an underlying theory of psychological processing, and the robustness of the accompanying task analysis. As the focus of this paper is on the interpretation of the neuroimaging data, I'm going to assume that the behavioural tasks used are reliable indicators of the modulation of attentional control. It is worth noting, however, that this assumption does not generally hold, especially given the relative lack of critical task analyses in neuroimaging research (see Poldrack 2010b for a discussion).

2.4.1 Deconvolution Analysis

Not all of the measured changes in the BOLD signal are relevant to the subject's performance of the cognitive task. The first substantive step in analyzing neuroimaging data is to extract the portion of the BOLD signal that corresponds with the task manipulation. This process is called deconvolution. Deconvolution is an algorithmic solution to a particular type of signal processing problem in which a signal of interest is convolved, or mixed with, another signal. In general, deconvolving the signal of interest requires solving an equation of this form:

$$(f \otimes g) = h$$

Where h is the recorded signal, f is the signal of interest and g is the signal that f needs to be separated from. In the case of fMRI data, h is the measured BOLD signal, g is the design matrix (a mathematical representation of the task) and f is the hemodynamic response function (hrf). The hrf represents the change in blood oxygenation levels that corresponds with the demands of the cognitive task that the subject performed. The aim of deconvolution analysis is to identify the portion of measured brain activity that is modulated by the task. Solving for the hrf requires pseudo-inverting the design matrix and multiplying it by the measured BOLD signal (this is the matrix-algebra equivalent of dividing both sides in the above equation by 'g' in order to calculate f).

It is important to note that this procedure only works when the trials are mathematically separable, which can be achieved using an event-related design. An event-related design is such that the stimuli or tasks are separated by an intertrial interval (usually there are about twenty seconds between tasks). Investigators can then assume that task-relevant BOLD activity occurs for short, discrete intervals corresponding to the onset of the task. The intertrial interval supports this assumption by ensuring that the trial-relevant signal is temporally localized, and does not uniformly influence subsequent trials.[8]

Mathematically, this amounts to assuming that task-relevant variation in the BOLD signal is linearly summed with the task-irrelevant BOLD signal, and so the two can be separated by the deconvolution procedure described above. It is worth noting that these (and the following) assumptions are supported by supplementary empirical research, and are not

---

[8] The intertrial interval does not need to be the same between every trial. Indeed, it is typically jittered, or randomly varied so that the interval between any two pairs of trials varies. The variation in intertrial interval is important for blocking certain confounds and artifacts that can arise when event onset is uniformly spaced. Since jittered events are still mathematically separable, I have omitted a detailed discussion of jitter for the sake of simplicity.

arbitrarily made or taken for granted (see Chapter 12 of Kass, Eden, Brown 2014 for a technical introduction to linear regression).

Typically, researchers assume that the hemodynamic response has a canonical shape and use that assumption to determine the form of the hrf. In this case, however, the investigators did not want to assume that the hemodynamic response function takes the canonical form and so they used a linear regression formula to model the hrf. This decision eliminates confounds that might arise from deviations in the hrf from the canonical model. The regression approach also allows the form of the hemodynamic response function to vary from voxel to voxel, instead of assuming that the BOLD signal follows the same pattern in every voxel.

Regression is a curve fitting procedure. The investigators specify an equation, a linear one in this case, with unknown coefficients, that is fit to the data. In this case, the 'data' that the curve is fit to is the result of multiplying the BOLD signal measurements with the inverted design matrix. The regression formula is expressed by the following equation: $x = \beta y + \epsilon$. Regression requires assuming that errors are independent (which is ensured by the event-related design) and that the noise term, ϵ, is linearly additive. For each regressor there will be an additional βy term. Liu and colleagues treated each experimental condition as a separate regressor, which resulted in a total of six regression terms (one for each of the clockwise, counterclockwise, red, green and null task conditions).

Once the regression formula and design matrix are determined, the design matrix is pseudo-inverted and multiplied by the measured BOLD signal. Then, the result of that is used to determine the unknown β values in the regression equation. Note that this procedure is implemented for each voxel, and so each voxel will have its own set of β values. The β values are then filled into the linear regression formula and the result is the hemodynamic response function.

The hemodynamic response function, as represented by the β values, indicates the portion of the measured BOLD activity that varies with task onset. This could be understood as capturing the portion of the data that is relevant to the manipulation of the

experiment. The β values are used in both the ROI selection procedure and the pattern classification analysis.

2.4.2 Region of Interest Selection

Once the hrf was calculated, the investigators used a goodness-of-fit measure to determine the amount of variance in the measured signal that the hrf accounted for. This provides an indicator of the portion of the signal that the hrf models accurately. To do this, they first averaged the modelled activity (the β values) over continuous groups of voxels (which they took to indicate specific regions of the brain). Then, they calculated the goodness-of-fit of the hrf, which is a measure of the amount of variance in the signal that is accounted for by the hrf. To evaluate the statistical significance of the estimate they used a permutation test (see Gardner et al. 2005 and Nichols and Holmes 2002 for details on these procedures).

Where the hrf identifies the portion of the signal modulated by the experimental tasks, the goodness-of-fit measure specifies the regions of the brain (understood as a collection of nearby voxels) where the hrf accounts for a significant portion of the variance in the BOLD signal data. The result of the procedure identifies regions of the brain where the variation in activity is correlated with the task demands of the experiment. When the variance of activity in a region accounted for by the hrf was sufficiently high, the investigators concluded that activity in that region '. . . is modulated by feature-based attention' (p. 4488).

This interpretation of the analysis result is a hemodynamic hypothesis since it relates variation in BOLD signal activity to specific task conditions. The particular hemodynamic hypothesis advanced attributes the portion of the measured BOLD signal captured by the β values that satisfy the goodness-of-fit criteria to the behavioural tasks. Calculating the hrf identifies the portion of the signal that corresponds with the onset of each task condition, eliminating the task-irrelevant portion of the signal. The goodness-of-fit procedure identifies the areas of the brain for which the hrf accounts for a significant portion of the variance in the activity. In other words, this ROI selection procedure identifies the regions in which the measured variation of the BOLD signal can

be explained in the context of the experiment. The result is used as a pre-processing step to select regions of interest for pattern classification analysis. As I will show, this step improves the strength of the experimental evidence for the theoretical hypothesis the investigators infer by providing partial validation for a crucial assumption implicit in the use of pattern classification analysis.

2.4.3 Pattern Classification Analysis

The primary aim of the study was to use pattern classification analysis to test '. . . whether the pattern of fMRI response across voxels in an area could distinguish which feature was attended, although the average amplitude[9] did not' (p. 4490). Pattern classification analysis is a type of multivariate analysis technique that treats each voxel as a dependent variable. The procedure involves four distinct stages: feature selection, classifier selection, training and testing. Feature selection involves choosing the voxels that will be included in the analysis. Typically, the chosen voxels are those within a particular ROI, although how that ROI is defined varies from study to study. Regions of interest can be defined anatomically, either using software to select the voxels that fall within the anatomical ROI, or by manually tracing the ROI. They can also be defined functionally, using a functional localization task. The BOLD signal data collected while a participant performs such a task can be used to identify voxels that are strongly activated during the performance of that task, which are then defined as the ROI. In this case, the investigators selected the voxels indicated by the procedure discussed in the previous section.[10]

---

[9] The investigators reported on a third analysis in the paper that I do not discuss in detail. That analysis, which adheres closely with the logic of subtraction, was intended to investigate if average BOLD signal amplitude discriminated between the specific features attended (red-dots vs. green-dots). It did not.

[10] In addition to the analyses I discuss in detail, they also completed a whole-brain searchlight analysis. A searchlight is a specific kind of feature selection and analysis

Classifier selection involves choosing the classifier, which is a machine-learning algorithm that will be used to implement the analysis. The classifier represents brain activity in a multidimensional space where each dimension corresponds to the BOLD signal value in each voxel. If three hundred voxels are selected, then the space has three hundred dimensions. Each point in this space specifies a particular BOLD signal value for each selected voxel and so corresponds to a particular state of brain activity. For the purposes of this paper, the particular classifier used does not matter, but it is worth noting that different classifiers have different strengths and weaknesses (see Misaki et al. 2010).

Once the classifier is selected it is trained and tested. During the training phase, the classifier is presented with labelled data (the labels indicate the task condition, such as 'attending to clockwise rotating dots'). The classifier identifies correlations between patterns in the BOLD signal and the provided labels, and based on those correlations it divides the multidimensional space into subspaces. Different classifiers use different procedures for subdividing the multidimensional space. Once subdivided, the classifier identifies each subspace with the task condition that is most frequently associated with it.

During testing, the classifier is presented with unlabelled data that it has not seen. It locates the novel data in the multidimensional space and, based on the subspace that it falls into, predicts the task label that corresponds with the data. A data point that is located in the 'attending to red' subspace is labelled as 'attending to red'. The predicted labels are compared with the true labels and the classifier's accuracy at predicting the task condition on the basis of the BOLD signal data is calculated. The regions of the

---

process. In a searchlight, investigators define a volume (the 'searchlight'), and then run the pattern classification analysis procedure over voxels within that volume. Then, they move the volume and run the analysis again. This procedure is typically used to identify arbitrary subdivisions of the brain that result in reliable classification, or to examine how classification accuracy changes as the classifier is given data from different parts of the same network or part of the brain.

brain (as defined by the ROI selection procedure) where the classifier performed with sufficient accuracy are said to '. . . contain the control signals for maintaining attention to visual features' (p. 4493). That is to say, the investigators took the classification results to indicate the regions of the brain that contain signals used for the maintenance of attention. They are attributing a cognitive function to a particular region of the brain (in fact, several regions of the brain). This is an inference to a theoretical hypothesis. In this case the hypothesis specifies the particular role that the identified regions perform - control of attentional processes.

The attribution of functional role is made on the basis of the information carried in the signal that is necessary to support the cognitive function. It's not just a claim that the indicated regions 'play such and such a role', but, by basing this inference on pattern classification analysis, it is a specification of that role in terms of the signal content. Given this, the inference from the successful predictions of a pattern classifier to the content of the brain activity, and subsequent attribution of functional role, requires additional assumptions. One particular assumption is that the patterns leveraged by the classifier contain information that is accessible to the brain.

One way to understand why this assumption is required is by distinguishing between the informational and representational content of a signal. The informational content of a signal is whatever facts you can learn from the signal. The representational content of the signal is the message actually carried by the signal. Informational content and representational content are not necessarily the same (Dretske 1981). Consider the following simple case: you are in a closed room and someone in an adjoining room is communicating a message by banging objects together. Perhaps they are using Morse code to express a fact about the weather. With sufficient equipment and expertise you could determine if the person in the other room is moving around, or features about the materials that they are banging together. These facts are part of the informational content of the signal, as they are facts you can learn by analyzing the signal. The actual message being communicated, however, may have nothing to do with these facts. Indeed, in this case the message is about the weather. It may even be the case that the individual who is communicating does not have access to the facts you are able to infer from the signal.

They may not know what material the objects are made of, and so could not possibly be communicating those facts. Without some knowledge of Morse code, or additional constraints beyond the signal itself, it is difficult to verify that facts learned from analyses of the signal correspond to the representational content of the signal. Thus, showing that regularities in a signal can be used to reliably make inferences or predictions about the world, as pattern classification analysis does, is not sufficient to support the claim that the signal is transmitting those facts.

In these terms, pattern classification analysis characterizes some of the informational content of the BOLD signal. It identifies which tasks can be discriminated between on the basis of patterns in the signal. The inference from the informational content of the BOLD signal to an attribution of functional role requires the assumption that the informational content extracted by the analysis reflects the representational content of the signal. Thus, successfully making an inference to the role a region plays on the basis of pattern classification requires, at least, that the information leveraged by the classifier is accessible to the brain or, more broadly, the organism.

Neuroscientists are well aware of this limitation. Classifiers are known to be very powerful and researchers caution against drawing inferences from the particular decision metric that a classifier implements. This is because a classifier will leverage anything that permits it to make reliable predictions, including patterns in the data irrelevant to understanding the functioning of the brain (Anderson and Oates 2010). Tong and Pratte relate an illuminating case of a classifier achieving near perfect accuracy at predicting the experience of humour when a subject was watching a sitcom while in an MRI scanner (2012). A close inspection of the classification process revealed that several voxels in the data were located along the edge of a ventricle (ventricles are a hollow space in the brain filled with cerebrospinal fluid). Since the ventricles contain no blood, the BOLD signal there is zero. Thus, a voxel along the edge of a ventricle will display a significant change in BOLD signal value should the subject's head move (even slightly), such as when stifling laughter. The classifier's performance was due to a correlation between slight head motion, humorous stimuli and voxels that overlap with ventricles. This is why researchers use secondary analyses, such as the ROI selection procedure described above

and the searchlight procedure described in footnote ten. These procedures help limit the possibility of the classifier 'cheating', which in turn provides (some) validation for the assumption that the information in the signal leveraged by the classifier is accessible to the brain.

2.5 The Strength of Multiple Analyses

The analysis techniques discussed above support different types of hypotheses. The ROI selection procedure supports a hemodynamic hypothesis about the relationship between variation in the BOLD signal and variation in the task conditions. Pattern classification analysis is taken to support a theoretical hypothesis about the functional role played by parts of the brain in attentional processes. The difference in use reflects a difference in evidence.

ROI selection identifies the portion of the data that can be explained in the context of the experiment. Pattern classification analysis identifies the task conditions that can be discriminated between on the basis of the fMRI data. The goodness-of-fit measure does not provide evidence that could support a claim about what task conditions can be discriminated between on the basis of the neuroimaging data. Likewise, the result of pattern classification analysis cannot support a claim about the quality of the data, or characterize which portion of the signal is modulated by the experimental manipulation. Indeed, that the classifier will leverage any correlation between task label and fMRI data suggests that it is particularly poorly suited to provide evidence in support of such a claim. The difference in evidence can be traced to a difference in the manipulations of the data. Through their different manipulations, the different techniques reveal different patterns.

Using these analyses together strengthens the evidence provided by classification analysis with respect to the target theoretical hypothesis. The permutation test indicates the portion of the signal that can be explained in the context of the experiment. By using the results of that procedure to select features for the classifier, the investigators ensured that the patterns available to the classifier are only those contained in the portion of the signal that is modulated by the experimental task. While this does not guarantee that the

leveraged signal carries information that is accessible to the system, it ensures that the leveraged variations are at least relevant to the experimental manipulation. In this way, some of the confounds that might prohibit inferring from the result of classification analysis to the target theoretical hypothesis are controlled for by using multiple analyses in series.[11]

The permutation test, when used to select a portion of the data for classification, provides validation for one of the problematic assumptions invoked by pattern classification analysis. Not only do these analysis techniques have different evidential targets, but brought together they provide stronger evidence for a theoretical hypothesis than either could alone. In this way, multiple analysis techniques that provide different perspectives on the same data and can strengthen the evidence produced in a single neuroimaging experiment. This is a kind of local robustness.

Robustness has been used to defend experimental practice from critiques similar to those discussed here. Specifically, Collins' experimenter's regress proposes a vicious circle between experimental results and the techniques that produce those results. He argues that a technique is verified only when it produces correct data, but a technique is only known to produce correct data when it is verified (1985). The critiques raised against neuroimaging by van Orden and Paap, which form the foundation of skepticism towards the technology, are of a similar form. The main issue they identify is that subtraction analysis requires assuming that the brain can be subdivided into functional parts, which is the very claim the analysis result is taken to support. This is a localized case of the experimenter's regress where the feature of scientific practice under scrutiny is not an instrument, but a data analysis technique.

Philosophers have argued that, with respect to the experimenter's regress, the epistemic situation is not as dire as Collins makes it out to be. Cartwright, for example, argues that

---

[11] Although not all. I leave discussion of those details for future work as it is beyond the scope of this paper.

the regress is broken by the robust reproducibility of instrument results (1991). Confidence in the report of an instrument is justified when the measurement result aligns with results produced by a variety of instruments, each of which relies on independent assumptions (p. 451-452). Culp offers a more careful defense along the same lines (1995). She argues, via a detailed case study analysis of approaches to DNA sequencing, that experimentalists are convinced that measurements are getting at the same phenomenon when multiple measurement techniques, each with different theoretical presuppositions, produces a robust body of evidence (p. 441).

Robustness is achieved when the same result is obtained by multiple, independent (or mostly independent) techniques (Wimsatt 1981). Robustness analysis involves determining the features of measurement or analysis techniques that are invariant under changes in the technique that might influence the result (Calcott 2011). Robustness is derived from the use of multiple independent approaches to detecting, isolating or measuring the same target. The independence of measurement results is characterized in terms of theoretical presuppositions required by the use of the instrument. These can also be understood as assumptions researchers must make about the production of the resulting data. Different instruments are independent insofar as they require different assumptions. The same can be said of different data analysis techniques.

Data analysis techniques, because of the manipulations they impose on data, require investigators to make assumptions about the result. These assumptions, if true, justify interpretations of the result of the data manipulation or analysis procedure. Different techniques, as used to support different hypotheses, require different assumptions. However, there is a relevant difference between using multiple data analysis techniques as I have described, and the use of multiple measuring instruments to detect the same phenomenon. The robustness of a measurement outcome is improved when independent techniques produce the same result. A defense of neuroimaging against van Orden and Paap's criticisms along these lines is offered by Roskies with her account of functional triangulation (2010b). Functional triangulation occurs when different analysis techniques produce the same result, and so generate a robust body of evidence. The situation I have described is different.

The techniques discussed above do not, and indeed cannot, provide the very same result. While the results of the analyses are not precisely the same, they are similarly aimed. The permutation test indicates the regions of the brain that may play a role in attentional processing, and the pattern classification analysis further clarifies that role. Thus, while they do not provide evidence in support of the very same hypothesis, the hypotheses they individually support are mutually supportive. The permutation test provides support for a hemodynamic hypothesis, and the subsequent analysis of the evidence revealed by that test using pattern classification analysis is brought to bear on a theoretical hypothesis. Insofar as this is a robust result, then, it might be regarded as a weakly robust result. Weak because the techniques do not have the same outcome.[12]

In general, different data analysis techniques provide different perspectives on the same data, and the use of multiple analysis techniques together can strengthen the quality of evidence produced by a particular method or instrument. This can result in evidence that can support inferences that may not be warranted by the result of a single analysis technique or data manipulation. In this way, multiple analysis techniques used in series can provide experimental results a kind of local robustness. It is 'local' because the techniques ultimately depend upon each other. While the different perspectives are not fully independent, because one analysis technique is used as a pre-processing step for a subsequently applied technique, they still contribute to the robustness of the inference because different techniques reveal (and suppress) different patterns and rely on different assumptions. Their differences are what contribute to the strengthening of the evidence.

The general lesson of the experimenter's regress is that problematic assumptions can arise in the context of experimentation. The general lesson of the appeals to robustness is

---

[12] This should not be cause for skepticism, at least not skepticism that is localized to the particular case of neuroimaging. There is reason to believe that any difference between measurement techniques can contribute to a difference in the phenomena probed by those techniques (see Sullivan 2009 for a discussion of this with respect to neurobiology). If this is true, then weak robustness is the norm for scientific knowledge.

that those assumptions can (sometimes) be validated by comparing different perspectives of the same subject. With respect to skepticism towards the use of neuroimaging data, I have argued that problematic assumptions, which arise from the use of particular analysis techniques, can be validated by using different data analysis techniques that require different assumptions. This provides the inference with a (weak) local robustness.

2.6 Conclusion

I have demonstrated that different data analysis techniques provide evidence for different phenomena and that multiple analysis techniques can be used together to improve the epistemic situation in neuroimaging research. Thus, the debate about the epistemic status of neuroimaging, which is framed in terms of the logic of subtraction, is at best an evaluation of the limitations of analysis techniques that depend upon that logic. Sweeping conclusions about the range of hypotheses that neuroimaging technology can and cannot be used to investigate are not supported by this literature.

The argument presented above provides grounds for a mild optimism with respect to neuroimaging technology. That is, optimism that it can be used to do more than provide evidence about hypotheses specifying the relationships between BOLD activity and task performance. I leave identifying what specific hypotheses and phenomena neuroimaging technology can be used to investigate for future work, as completing this task will require a careful evaluation of a representative collection of the data analysis techniques and experimental strategies used in neuroimaging research. Given that different analysis techniques provide different evidence, the diversity of techniques used in neuroimaging research suggests that philosophers concerned with the epistemology of neuroimaging should focus their attention on evaluating the evidential quality and scope of particular analysis techniques (such as subtraction) and classes of analysis techniques (such as multivariate analyses). Such evaluations should take into account the specific theoretical goals they are put towards (functional localization, or tracking the content of neural representations, to name two).

The general lesson here is that data analysis techniques play an important role in the generation of scientific evidence. Differences in the data analysis procedure used and

differences in how that procedure is implemented can make a difference to the range of phenomena that the result of the analysis is informative about. This is a feature of scientific practice in need of more careful philosophical attention.

References

Aktunç, E. M. [2014]: 'Severe Tests in Neuroimaging: What We Can Learn and How We Can Learn It', Philosophy of Science, 81, pp. 961-73.

Anderson, M. L. and Oates, T. [2010]: 'A critique of multi-voxel pattern analysis', Proceedings of the 32nd Annual Conference of the Cognitive Science Society, pp. 1511-16

Ashby, F. G. [2011]: Statistical Analysis of fMRI Data. The MIT Press.

Bogen, J. and Woodward J. [1988]: 'Saving the Phenomena', Philosophical Review, 97, pp. 303-52.

Calcott, B. [2011]: 'Wimsatt and the robustness family: Review of Wimsatt's Re-engineering Philosophy for Limited Beings', Biology and Philosophy, 26, pp. 281-93.

Cartwright, N. [1991]: 'Replicability, Reproducibility, and Robustness: Comments on Harry Collins', History of Political Economy, 23, pp. 143-55.

Collins, H. [1985]: Changing Order. London: SAGE Publications.

Culp, S. [1995]: 'Objectivity in Experimental Inquiry: Breaking Data-Technique Circles', Philosophy of Science, 62, pp. 430-50.

Davis, T. and Poldrack, R. A. [2013]: 'Measuring neural representations with fMRI: practices and pitfalls', Annals of the New York Academy of Sciences, 1296, pp. 108-34.

Dretske, F. [1981]: Knowledge and the Flow of Information. The MIT Press.

Friston, K. J., Holmes, A. P., Price, C. J., Büchel, C., Worsley, K. J. [1999]: 'Multisubject fMRI Studies with Conjunction Analyses', NeuroImage, 10, pp. 385-96.

Gardner, J. L., Sun, P., Waggoner, R.A., Ueno, K., Tanaka, K., Cheng, K. 2005: 'Contrast adaptation and representation in human early visual cortex', Neuron, 47, pp. 607– 20.

Good, I. J. [1983]: 'The Philosophy of Exploratory Data Analysis', Philosophy of Science, 50, pp. 283-95.

Hardcastle, V. G. and Stewart, M. C. [2002]: 'What do brain data really show?', Philosophy of Science, 69, pp. S72-82.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai A., Scouten, J. L., and Pietrini, P. [2001]: 'Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex', Science, 293, pp. 2425-30.

Huettel, S. A., Song. A. W., and McCarthy, G. [2008]: Functional Magnetic Resonance Imaging. 2nd ed. Sunderland, MA: Sinauer.

Kass, R. E., Eden, U., and Brown, E. [2014]: Analysis of Neural Data. Springer-Verlag: New York.

Klein, C. [2010a]: 'Images are not the evidence in Neuroimaging', British Journal for the Philosophy of Science, 61, pp. 265-78.

Klein, C. [2010b]: 'Philosophical issues in neuroimaging', Philosophy Compass, 5, pp. 186-98.

Kriegeskorte, N. and Kievit, R. A. [2013]: 'Representational geometry: integrating cognition, computation, and the brain', Trends in Cognitive Sciences, 17, pp. 401-12.

Landreth, A. and Richardson, R. C. [2004]: 'Localization and the new phrenology: A review essay on William Uttal's the new phrenology', Philosophical Psychology, 17, pp. 107-23.

Liu, T., Hospadaruk, L., Zhu, D. C., and Gardner, J. L. [2011]: 'Feature-Specific Attentional Priority Signals in Human Cortex', The Journal of Neuroscience, 31, pp. 4484-95.

Logothetis, N. K. [2008]: 'What we can do and what we cannot do with fMRI', Nature, 453, pp. 869-78.

Martin, C. B., McLean, D. A., O'Neil, E. B., and Köhler S. [2013]: 'Distinct Familiarity-Based Response Patterns for Faces and Buildings in Perirhinal and Parahippocampal Cortex', The Journal of Neuroscience, 33, pp. 10915-23.

Mayo, D. [2005]: 'Evidence as Passing Severe Tests: Highly Probably versus Highly Probed Hypotheses', In P. Achinstein (ed), Scientific Evidence: Philosophical Theories and Applications, Baltimore: John Hopkins University Press, pp. 95-127.

McIntosh, A., Lobaugh, N., Cabeza, R., Bookstein, F., and Houle, S. [1998]: 'Convergence of neural systems processing stimulus associations and coordinating motor responses', Cerebral Cortex, 8, pp. 648–59.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., and Grady, C. L. [1996]: 'Spatial pattern analysis of functional brain images using Partial Least Squares', Neuroimage, 3, 143–157.

Misaki, M., Kim, Y., Bandettini, P. A., and Kriegeskorte, N. [2010]: 'Comparison of multivariate classifiers and response normalizations for pattern- information fMRI', NeuroImage, 53, pp. 103-18.

Nichols T. E., and Holmes A. P. [2002]: 'Nonparametric permutation tests for functional neuroimaging: a primer with examples', Human Brain Mapping, 15, pp. 1–25.

Poldrack, R. [2010a]: 'Subtraction and Beyond: The Logic of Experimental Designs for Neuroimaging', in M. Bunzl and S. J. Hanson (eds), Foundational Issues in Human Brain Mapping, The MIT Press, pp. 147-60.

Poldrack, R. [2010b]: ‘Mapping mental function to brain structure: How can cognitive neuroimaging succeed?’, Perspectives on Psychological Science, 5, pp. 753-61.

Roskies, A. [2010a]: ‘Neuroimaging and Inferential Distance: The Perils of Pictures’, in M. Bunzl and S. J. Hanson (eds), Foundational Issues in Human Brain Mapping, The MIT Press, pp. 195-216.

Roskies, A. [2010b]: ‘Saving Subtraction: A reply to Van Orden and Paap’, British Journal for the Philosophy of Science, 61, pp. 635-65.

Sullivan, J. [2009]: ‘The multiplicity of experimental protocols: A challenge to reductionist and non-reductionist models of the unity of neuroscience’, Synthese, 167, pp. 511-39.

Tong, F., and Pratte, M. S. [2012]: ‘Decoding Patterns of Human Brain Activity’, Annual Review of Psychology, 63, pp. 483-509.

Uttal, W. [2001]: The New Phrenology, The MIT Press.

Uttal, W. [2011]: Mind and Brain: A Critical Appraisal of Cognitive Neuroscience, The MIT Press.

van Orden, G. C., and Paap, K. R. [1997]: ‘Functional Neuroimages Fail to Discover Pieces of Mind in Parts of the Brain’, Philosophy of Science, 64, pp. S85-94.

Wimsatt, W. [1981]: ‘Robustness, Reliability, and Overdetermination”, in Re-Engineering Philosophy for Limited Beings, pp. 43-74.

Chapter 3

3 The Interpretation of Neuroimaging Data as Explanations of Data Patterns

3.1 Introduction

Debates about the inferences neuroscientists make on the basis of neuroimaging data have persisted in the philosophical literature amidst significant changes in the methods and techniques used in the field (e.g., van Orden and Paap 1997; Uttal 2001; Klein 2010). In addition to their skeptical conclusions, these arguments are similar in a number of respects. The neuroimaging data discussed consists in measurements of blood oxygenation levels and is used to investigate cognitively relevant neural activity. The indirect relationship between the measured data points and the targets of inference is taken by some skeptics to be sufficient grounds for sweeping skepticism about the status of these claims (e.g., Aktunç 2014a). These arguments also share an approach to evaluating the inferential limitations of neuroimaging data. Skeptics tend to examine the evidential relationship between the results of data analysis techniques, such as subtraction and more recently pattern classification analysis (PCA), and the phenomena that neuroscientists deploy those techniques to investigate. That is the localization of cognitive functions and claims about neural representations respectively. The analyses provided by the skeptics make explicit the logic implicit in the use of the techniques, and argue that treating the results of subtraction as evidence for localization claims (Hardcastle and Stewart 2002), or PCA results as evidence for claims about neural representations (Ritchie, Kaplan and Klein forthcoming), is viciously circular, or otherwise invalid.

Debates about the claims neuroimaging data can provide evidence for, and the best methods for pursuing this research also occur between cognitive neuroscientists. For instance, multivariate pattern analysis (MVPA) is a relatively new collection of analysis techniques that are often used to evaluate what information is carried in, or represented

by, neural activity (Tong and Pratte 2012; Haxby 2010). Cognitive scientists have criticized this use of MVPA methods by pointing out that changes in blood oxygenation are indirectly related to neural activity, limiting the strength of inferences that can be made (de-Wit et al 2015), and by clarifying how variations in the analysis process contribute to the obtained results (Anderson and Oates 2010; Misaki et al 2010). These concerns parallel those raised by skeptics. They draw out inferential challenges that follow from the indirectness of the data, and examine how the particulars of the data analysis processes used to investigate specific claims and hypotheses should influence the inferences made on the basis of the analysis results. They diverge from the skeptical arguments in their conclusions.

Where skeptics often recommend conservative interpretations of the data, insisting for example that inferences should be restricted to claims about the relationship between behaviour and blood oxygenation levels (Aktunç 2014a), neuroscientists continue to use these data to make claims about the relationship between the brain and cognition. Indeed, there has been a steady increase in the use of MVPA and claims about representations in the brain. This is at least in part because MVPA techniques are regarded as useful for investigating new questions and hypotheses, including questions about the content and structure of representations in the brain, but also because neuroscientists continue to work to improve their use (see Haxby et al 2014). While there are debates about the inferential limitations of these techniques as noted above, most of the contributions to these discussions are made by investigators who rely on them in their own research. Neuroscientists and skeptics disagree about the implications of well-known challenges with the use of neuroimaging data as evidence for particular kinds of claims. Other contributions to the philosophical literature suggest two reasons for this disagreement: (1) skeptical accounts overlook epistemically relevant details of the practice, and (2) some philosophical approaches to evidence in science do not capture what is epistemically good about the results of data analysis techniques as applied to neuroimaging data.

Like the skeptical arguments, the defences of neuroimaging research that resist them are similar. They each point out that the critics have, in some manner or another, mischaracterized the research process. This includes (a) focussing on rarely used

statistical methods (Machery 2014 on Klein 2010), (b) simplifying the analysis processes considered (Landreth and Richardson 2004 on Uttal 2001), (c) failing to consider supplementary evidence (Roskies 2010 on van Orden and Paap 1997), and/or (d) overlooking the diversity of analysis processes used to evaluate neuroimaging data (Wright forthcoming on Aktunç 2014a). The general lesson here is that, while there may be value in explicating the inferential limitations of specific analysis techniques, it is important that these limitations are not taken to reflect limitations of neuroimaging research as a whole, or even the limitations of a given neuroimaging data set. No matter how central they appear to be to an empirical investigation, even powerful data analysis procedures like pattern classification analysis do not comprise the full evidential base for inferences in neuroimaging research (see Roskies 2010 and Wright forthcoming in particular).

Separately from the debate over the efficacy of neuroimaging research, Jim Bogen uses inferences in neuroimaging research to argue against the adequacy of Jim Woodward (2000) and Deborah Mayo's (1996) accounts of evidence in experimental science. This argument shows that some philosophical approaches to examining inferences in neuroimaging research are ill suited for addressing questions about the evidential value of the data with respect to the claims neuroscientists use it to support. Mayo and Woodward's views locate evidential value in the ability to account for errors and produce data that are accurate with respect to the content of the claims the data are used as evidence for. Identifying, eliminating and reducing the impact of errors is certainly an important part of the interpretation of neuroimaging data. This is the role that many data analysis techniques perform, including the tools of inferential statistics and the data manipulations used to eliminate artifacts, reduce noise, and amplify signal. However, this is not the only contribution that data analysis techniques make to this process, and treating it as such leads to systematically undervaluing neuroimaging data. Indeed, Mayo's account is at the heart of a recent contribution to the skeptical tradition that argues that neuroimaging data's value is fundamentally limited because the techniques used to analyze it, such as subtraction, do not constitute tests with sufficiently stringent error-characteristics (Aktunç 2014a; 2014b).

Orthogonal to the skeptical literature, Bogen argues that this should not be taken to reflect a failure of inferences in neuroimaging research, but instead taken as a problem for Mayo and Woodward's accounts of evidence. The problem, Bogen argues, is that neither account fully captures the evidential value of neuroimaging data (Bogen 2002). His argument turns on the observation that "... what is epistemically good about functional images is not that they are highly accurate with regards to [biological indicator] levels or locations of individual brains" (p. S65), and further that the purpose of manipulating imaging data is not to "... bring error-ridden, [biological indicator] estimates recognizably closer to what would have resulted from ideal experiments shielded from significant sources of error" (p. S65). Bogen's proposal is that we should "… think of functional images as perfectly accurate, error free representations of the anatomy and cognitively significant [biological indicator] of highly idealized, imaginary brains" (p. S67). On this view, the epistemic value of the functional image is determined by the degrees of resemblance between the idealized brain it portrays and the real brains it is used to make claims about. The images need not be accurate in every respect, but only in those respects that are relevant given the hypotheses under investigation (p. S68-9).

While Bogen focusses on neuroimages, which are one product of data analysis in neuroimaging research, a similar argument could be made for MVPA techniques like pattern classification analysis. The purpose of these techniques is not to better approximate the results of an ideal experiment — such as one directly measuring the representationally-relevant dimensions of neural activity — and so analyses of the evidential value of neuroimaging data cannot treat them as such. The problem with the skeptical literature is not just that skeptics ought to include more details about the practice in their analysis, but that their approach to evaluating the inferences made in neuroimaging research may itself need revision. The aim of this paper is to argue for a new approach to examining inferences in neuroimaging research that is sensitive to the epistemic role data analysis techniques actually play in those inferences.

My first step towards this is methodological. Instead of examining the logical structure of particular inferences, I examine how data analysis techniques are used to arrive at those

inferences in the first place. In doing so I focus on the data interpretation process in which scientists use data analysis techniques to assess the evidential relations that hold between the data at hand and a claim, and not the product of that process — which is an inference, or judgement of evidential value. This work is informed by my experience collaborating with neuroscientists, both as a lab member and as a collaborator (e.g., Martin et al 2015). This experience is what made clear to me that analyses of the structure of inferences from data to claims must be complemented by an examination of the processes and assumptions that scientists use to navigate and engage with the challenges that complicate those inferences.

Inspired by Bogen's view on the epistemic value of neuroimages, in what follows I show how data patterns are in general inferentially limited and yet valuable as evidence for the claims neuroscientists regularly infer. A conceptual framework that is adequate with respect to the actual practice of data interpretation in neuroscience must provide an account of how the limited evidence provided by data patterns could be accumulated into evidence sufficiently strong to justify the claims inferred. Julian Reiss's pragmatic account of evidence provides a good starting point for articulating the aspects of this process (2015). In applying it I argue that the account of evidential accumulation Reiss proposes is inadequate in the case of neuroimaging research. I argue instead that this process has an explanatory character, and appeal to I.J. Good's work on exploratory data analysis to draw a parallel between the process of interpreting neuroimaging data and the processes of exploratory data analysis (1986).

I proceed as follows: in section 2, I provide an overview of neuroimaging experiments and outline three challenges with using neuroimaging data as evidence in cognitive neuroscience. In doing so I identify an apparent tension in the use of data analysis techniques to assess the evidential value of neuroimaging data. In section 3, I diffuse this tension by looking more closely at pattern classification analysis (PCA) and the relationship between its results and claims it is used to investigate. By distinguishing between claims about data, claims about phenomena, and the evidential relations that can hold between data and such claims, I show how the inferential limitations of data analysis techniques outlined in section 2 do not undermine their usefulness in assessing the

evidential value of neuroimaging data. In section 4, I argue that the strength of evidence that a data set provides for a claim is a function of the explanatory relation that holds between the claim and each of a variety of data patterns isolated by different analysis techniques or variations of a single technique. I conclude in section 5 with reflections on further questions raised by this account.

3.2 Neuroimaging Experiments

Neuroimaging experiments[13] are used to learn about the relationship between brain activity and cognitive functioning in humans. The research that I will focus on involves the use of functional magnetic resonance imaging (fMRI) to measure changes in the ratio of oxygenated to deoxygenated hemoglobin in the brain, called the BOLD (blood oxygenation level dependent) signal, as a subject performs a variety of cognitive tasks (see Ashby 2011 for an introduction). Experiments involve placing a human subject in a magnetic resonance imaging scanner while they perform a cognitive task, such as attending to a rotating pattern of dots (Liu et al 2011), or deciding whether or not an image is one they saw in a previous part of the experiment (Martin et al 2013). While the participant performs the task, the scanner, using a functional scanning protocol, measures changes in blood oxygenation throughout their brain. The resulting data set is used as evidence for claims about the relationship between brain activity and cognitive processing. This includes claims about the role regions play in cognition, such as the claim that the hippocampus is involved in memory (e.g., Greicius et al 2003), and claims about the information carried or represented in neural signals, such as claims that the content of memories can influence their maintenance (e.g., Kim et al 2014). The first

---

[13] The discussion of experimental practice I provide is only as detailed as required for the discussion that follows. It is informed by a variety of accounts of experimental practice (Hacking 1983; Bogen and Woodward 1988; Mayo 1996; Woodward 2000; Sullivan 2009), and accounts of scientific data (Hacking 1983; Rheinberger 2011; Leonelli 2015). For broader perspectives on experimentation in science Hans Radder's volume *The Philosophy of Scientific Experimentation* (2003) is an excellent place to start.

claim is an example of a ‘localization claim’, since it attributes a cognitive role to a particular region, network or part of the brain. The second claim is an example of what I will call a ‘representational claim’, since it is about the information carried in, or represented by, patterns of brain activity (in this case, the content of memories).

Skeptics tend to emphasize two challenges for these inferences: (1) neuroimaging data consists in measures of blood oxygenation and behaviour, and thus cannot be used to reliably make inferences about neural activity and cognitive processes (Aktunç 2014a; 2014b); and (2) the techniques used to analyze neuroimaging data require investigators to make assumptions that undermine the inferences those techniques are used to support (Hardcastle and Stewart 2002; Ritchie, Kaplan and Klein forthcoming). The first of these challenges is often supported by an argument from underdetermination. The indirect nature of the data, combined with the complex interconnectedness of neural systems, ensures that there are a large number of viable alternative claims that the results of a neuroimaging experiment are consistent with. This is regarded as sufficient grounds for suspicion (Mole and Klein 2010; Klein 2012; Aktunç 2014a). The second challenge usually accompanies a logical analysis of the use of a popular data analysis technique, such as subtraction or pattern classification analysis. These analyses identify assumptions implicit in the operation of the techniques, or the most common interpretations of their results, that undermine the inferences either because the assumptions are likely to be false, or worse make the inference viciously circular (van Orden and Paap 1997; Uttal 2001; Hardcastle and Stewart 2002; Ritchie, Kaplan and Klein forthcoming).

As noted above, arguments against these skeptics tend to identify epistemically relevant aspects of the data interpretation process that skeptics deemphasize, overlook or fail to take into account. While these are grounds for denying the conclusions, skeptics arrive at, the skeptical views are not without merit. Even if their conclusions are not warranted because they overlook a relevant aspect of the practice they criticize, they still identify genuine inferential challenges characteristic of neuroimaging research. These are the challenges that investigators must somehow overcome if they are to justifiably infer, for example, representational claims, on the basis of data produced in a neuroimaging experiment. An adequate account of the interpretation of neuroimaging data must identify

how these challenges originate, and how the interpretation process engages them. The aim of this section is to address the first requirement, while the remainder of the paper addresses the second.

I have divided this section into two parts: data production and data interpretation (following Woodward 2000). The primary reason for this is pragmatic, as doing so makes it easier to locate the origins of the challenges noted above and identify a tension in the use of data analysis techniques to address them. To this end, I use data production to refer to any and all processes by which investigators assemble, produce or access a collection of data for the purpose of learning about one or more target phenomena. Data interpretation, on the other hand, refers to the processes investigators engage in to assess the evidential value of the produced data with respect to claims about the phenomena of interest. The product of data interpretation is an inference from the produced data to one or more claims about phenomena.

3.2.1 Data Production

Data production in neuroimaging research involves collecting data about structural features of participant's brains, changes in brain activity, and details about each participant's behaviour during the experiment. This comes in the form of three distinct data sets: structural, functional and task data. These data sets are used together to make claims about the features and anatomical origins of cognitively relevant brain activity and the cognitive processes, states, functions or capacities that are correlated with that activity.

Structural data are obtained using a magnetic resonance imaging scan. This data set captures information about anatomical features of the scanned brain such as the shape of regions, the size and location of folds (sulci and gyri), and the division between grey matter (areas mostly consisting of neuron cell bodies) and white matter (areas mostly consisting of axons).

Functional data are collected using scanning protocols that measure the ratio of oxygenated to deoxygenated hemoglobin in small 'pieces' of the brain. The pieces that

the scanner divides the brain into are called voxels, which are typically 2 to 3 mm cubes.[14] A functional scan measures the ratio of oxygenated to deoxygenated hemoglobin within each voxel. This is called the BOLD, or blood oxygenation level dependent, signal.

Task data consists in features of the task, stimuli parameters, and measurements of task performance. Task data are used to identify the cognitive functions, states and processes engaged by the participant during the experiment. This data consists in measurements of externally accessible behaviours that are used to make claims about cognitive processes internal to the subject that are themselves not directly detectable. In this way, the data are indirectly related to the phenomena that they are used to make claims about. Steps are taken in the production stage to overcome this limitation and provide opportunities for addressing it during the interpretation stage.

Investigators design tasks to target specific cognitive states or processes and use behavioural measures to detect instances of those states or processes. For example, in an experiment where participants must attend to one of two overlapping gratings, the investigators may require participants to press a button when the grating they are attending to changes in size. By controlling when each grating changes in size the behavioural data can be used to accurately classify each instance of the task (Kamitani and Tong 2005). Task parameters are also used to discriminate between instances of cognitive processes. For example, the properties of a field of dots can be varied in order to contrast attention-to-colour with attention-to-motion (Liu et al 2011). Typically, a

---

[14] The relatively new 7-Tesla scanners have more powerful magnets than the 3.5-Tesla scanners that have become the standard for neuroimaging research. 7T scanners, since the increase in field strength increases the signal, allow for functional scans with voxels as small as 0.7mm cubes. The increase in signal, however, calls for new methods to analyze the data. Work is ongoing to understand how 7T measurements compare with the results of weaker scanners, and for the sake of simplicity I will not discuss them here (e.g., Seiger et al 2015).

combination of behavioural measures, properties of the stimuli, and task demands are used to control, and later evaluate, what cognitive processes are engaged by subjects during an experiment.

Like task data, functional data are also indirectly related to the causal factors that give rise to the phenomena they are used to investigate. This is one of the central challenges with using neuroimaging technology to investigate the relationship between cognitive and neural processes. The fact that access to information about neural activity is mediated through measures of the BOLD signal places limits on the specificity of the claims that BOLD signal data can support about neural activity. For instance, it is known that "[t]he fMRI signal cannot easily differentiate between function-specific processing and neuromodulation, between bottom-up and top-down signals, and it may potentially confuse excitation with inhibition" (Logothetis 2008, p. 877). That is to say, there is a number of functionally distinct neural activation patterns that can give rise to a given BOLD signal measurement. This fact is often noted by critics of inferences made on the basis of neuroimaging data (e.g., Klein 2012; Aktunç 2014a).

Compounding the challenge arising from the indirect relationship between the data and phenomena is the growing body of evidence showing that observed localized change in brain activity, say in the hippocampus, could be the result of any of a number of possible upstream activities. This is because a number of the brain regions have connections that feed into the hippocampus, and any of those — or a combination of them — could be responsible for the observed change. This has been identified as an inferential challenge by skeptics (Klein 2012), and neuroscientists (Marder 2015). This adds to the number of claims that a given result might support, further amplifying the underdetermination of inferences in neuroimaging research that was already high due to the indirect nature of the data.

A well-designed experiment is one that is capable of producing data that has the potential to recommend accepting or rejecting the hypothesis it was produced to investigate. However, a data set does not, merely by its production, provide evidence for or against any particular hypothesis or claim. When the data is in hand investigators must make a

judgement about which claims the data provides evidence in favour of, and which claims they provide evidence against. To make such an assessment is to make a judgement of the evidential value of the data with respect to a number of claims about phenomena. Thus begins the stage of data interpretation.

3.2.2 Data Interpretation

Assessing the evidential value of neuroimaging data requires investigators to analyze and manipulate it. Scientists working with neuroimaging data are well aware of the limitations discussed above, that is the indirect nature of the data and the implications of the interconnectedness of brain systems. To determine whether or not a given data set provides evidence for or against a target claim, investigators manipulate the data. While there are data manipulations, referred to as 'pre-processing steps', whose primary function is to eliminate errors in the data (see Poldrack, Mumford & Nichols 2011, Ch 3, p. 34-50), the techniques skeptics emphasize in their critiques are not amongst these procedures. Techniques like subtraction, and pattern classification analysis, perform a distinct epistemic role. Instead of correcting for errors, they are used to directly assess the evidential value of the data set with respect to specific claims. Localization claims in the case of subtraction, and claims about representations in the case of pattern classification analysis.

Take pattern classification analysis (PCA) as an example. PCA involves training a machine learning classifier to correlate BOLD signal measurements with either task conditions (such as 'dots rotating clockwise' or 'dots rotating counterclockwise'), or cognitive processes (such as 'attending to rotating dots' or 'attending to counter-clockwise rotating dots'). Then, the classifier is provided with BOLD signal data and assigns a label to it. When the classifier's accuracy at labelling the functional data is significantly above chance investigators interpret this as showing that information in the BOLD signal permits the discrimination of the task conditions or cognitive processes the classifier was trained to label (as in Haxby et al 2001; Kamitani and Tong 2005; Liu et al 2011; Martin et al 2013; Sandberg et al 2014). PCA results are often treated as evidence for claims about information in or carried by brain activity (Norman et al 2006), an inference that philosophers have recently criticized for assuming that the ability to decode

a signal is directly informative about the information carried in that signal (Ritchie, Kaplan and Klein forthcoming). Concerns about what can be inferred from the results of PCA are echoed in the scientific literature (Anderson and Oates 2010; Davis and Poldrack 2013; de-Wit et al 2015).

This points towards a tension in the use of data analysis techniques to find patterns in the data that are informative about specific hypotheses or phenomena. On one hand, tools like PCA are necessary for assessing the evidential value of neuroimaging data with respect to claims about phenomena that are indirectly related to the measured data points — representational claims in particular. Without techniques for identifying patterns in the data that are informative about the target phenomenon, neuroimaging data are at best useful for studying the relationship between blood oxygenation levels and behaviour. On the other, the use of these techniques introduces assumptions into the inference from the data to the target claim — assumptions that have been identified as threats to the validity of those inferences (van Orden and Paap 1997; Hardcastle and Stewart 2002; Anderson and Oates 2010; Davis and Poldrack 2013; Ritchie, Kaplan and Klein forthcoming).

Even with knowledge of these challenges, neuroscientists continue to regard neuroimaging data as a valuable source of evidence for claims about the informational content of brain activity, and the relationship between the brain and cognition more generally (Haxby et al 2014). In the rest of this paper I examine the process of data interpretation neuroscientists typically engage in, and identify how that process is sensitive to these challenges. Data analysis techniques like PCA play an important role in this process, and accounting for their contribution requires diffusing the apparent tension in the use of data analysis techniques to interpret neuroimaging data. This is the task of the next section.

3.3 Data and Evidence

The indirect relationship between the data and phenomena they are used to investigate make it necessary to manipulate the data in order to assess their evidential value. However, the techniques used to do so appear to rely on, or invoke, assumptions that can undermine the resulting inferences to claims about the target phenomenon. In this section

I diffuse this apparent tension. Doing so requires distinguishing between kinds of claims about phenomena that arise as part of the data interpretation process, and also distinguishing between two evidential relations that can hold between evidence and a claim. Making these distinctions shows how techniques like pattern classification analysis are unable to provide sufficient evidence to justifiably infer representational claims, and yet can be useful for assessing the evidential value of the data they are derived from with respect to such claims. I begin with an overview of pattern classification analysis.

Pattern classification analysis (PCA) is a multivariate analysis technique that is used to examine the informational content of brain activity (Norman et al 2006; Haxby 2014). The results of PCA indicate the extent to which it is possible to use one set of experimental variables to predict the values of another. The procedure of implementing PCA has four major steps: (1) feature selection, (2) classifier selection, (3) training, and (4) testing.

Feature selection involves identifying a portion of the functional data set to be used in the pattern classification analysis. Feature selection is driven by considerations of the hypothesis or claim of interest, and ideally picks out data points that are most likely to be informative about that target. Features may be selected by their anatomical location (Liu et al 2011), by their responsiveness to particular task conditions (Martin et al 2013), a mixture of these factors, or by a stochastic procedure (Etzel et al 2013). Classifier selection involves selecting, or programming, the machine learning classifier that will be used in the analysis. Both pragmatic considerations and features of the hypotheses and data sets are relevant for selecting a classifier. For instance, a Gaussian Naive Bayes classifier is better for procedures that need to be repeated may times since it can be trained faster than others, while Support Vector Machines tend to work best when there are only two experimental conditions that the classifier has to select between (Pereira, Mitchell and Botvinick 2009; Pereira and Botvinick 2011).

With features and a classifier in hand, the subset of data picked out by the feature selection process is divided into two portions, one for training and one for testing. The

functional data in the training set are provided to the classifier along with labels of the task or cognitive process it is associated with. During training the classifier develops a decision procedure, based on correlations between BOLD signal patterns and the labels that it will use to label novel data. Finally, the labels are removed from the portion of the data that it was not trained on and the classifier assigns labels to these data according to its decision procedure. The assigned labels are compared against the true labels and the accuracy of the classifier is calculated. The accuracy of the classifier during the testing phase is the primary output of this analysis procedure.

The primary advantage cited for using PCA, and MVPA methods more generally is that they can be used to 'decode' the neural signals (Norman et al 2006; Kriegeskorte, Mur and Bandettini 2008; Tong and Pratte 2012). Some advocates of the method go so far as to claim that MVPA techniques have "… allowed researchers to access the contents of thoughts in considerable detail" (Haynes 2012, p. 30). Others recommend restraint in treating the results of PCA and similar techniques as 'reading off' the neural code (Anderson and Oates 2010; Davis and Poldrack 2013). Even the cautious, however, still regard the technique as useful for investigating claims about the content of representations in the brain, at least when the methods of analysis and interpretation are appropriately tempered by considerations of their limitations (Etzel et al 2013; Davis et al 2014; Haxby et al 2014).

MVPA techniques like PCA are often advocated for on the grounds that they bring "... fMRI investigation closer to investigating the codes for how functions are represented in neural population responses…" (Haxby 2010, p. 56). PCA techniques allow for a 'closer look' insofar as they are sensitive to smaller variations in the BOLD signal than other methods of analysis, not because they 'pick out' the representations directly (Haxby 2010, p. 57). MVPA methods, of which PCA is an example, are sensitive to different kinds of variations in the data than the subtraction methods that originally dominated the research literature. In particular, they are sensitive to multidimensional effects that cannot be detected by techniques like subtraction. This increased sensitivity comes at a cost, as there are some patterns subtraction and other univariate methods are better suited for highlighting (Davis and Poldrack 2013).

Inferring, on the basis of successful classification, a claim about the content of a representation in the brain, has been a recent target of criticism in the philosophical literature (Ritchie, Kaplan and Klein forthcoming). The interpretation objected to is treating significant classifier accuracy as "… strong evidence that the information is represented by the patterns of activity used as the basis for the decoding" (p. 8). Such an inference starts from a successful classification result, which establishes that the functional data contains sufficient information to predict the task labels, and from there concludes that this information is available to the classifier because it is represented in the brain activity underlying the BOLD signal measurements.

This inference, the authors argue, is undermined by a 'fundamental methodological issue' with classification techniques (p. 9). The problem is in assuming that the classifier uses the same information carried by, encoded in, or represented by the neural activity underlying the BOLD signal to label the novel data. However, classifiers are known to rely on any available correlations to make their predictions, and their performance is influenced by a number of decisions made in the course of the analysis. For example, the particular classifier (Misaki et al 2010), the method used to select data points for classification (Pereira, Mitchell and Botvinick 2009), and the timing windows that the functional data are divided into (Kohler et al 2013), can all influence classification accuracy.

It is on these grounds that Ritchie and colleagues conclude that "[a]t best, MVPA-based decoding shows that information about experimental conditions is latent in neural patterns. It cannot show that this information is used, or is even usable, by the brain" (p. 15-6). In a footnote, they argue that this argument retains its force even if the inference were adjusted, and classification results are taken to be weak (as opposed to strong) evidence for claims about representations (p. 16). Classification accuracy is insufficient, on its own, to warrant any claim about representations.

This exemplifies the tension noted in the previous section: PCA is used because it is regarded as informative about the content of representations in the brain, but the way the analysis technique is applied and the decisions made in its implementation conspire to

undermine its value as evidence for claims about representations in the brain. If the function of these techniques is to 'close the gap' between the data and target phenomenon by approximating the results of measurements of the relevant casual factors, then it would be fair to treat the classification results as the evidence upon which claims about representations are inferred. This account of their epistemic role mischaracterizes the contribution analysis techniques like PCA make to assessments of the evidential value of the data.

PCA allows investigators to determine if a task or cognitive processes can be predicted or classified via variations in the functional data[15]. Even considering the assumptions and challenges noted above, accurate classification is informative insofar as it demonstrates that information is available in the functional data. While PCA results cannot justify an inference to a claim about representations, they can be used to justifiably infer a claim that 'pattern-variations in the BOLD signal permit reliable discrimination of task conditions X, Y and Z', which is a claim about what the classifier has shown to be possible. This inference is not disputed by Ritchie and colleagues. They object to the further inference from these results to a claim about representations. The first step towards diffusing the tension in this use of data analysis techniques is to distinguish between the kinds of claims data analysis results can, on their own, justify and those they cannot.

The assumptions involved in PCA prohibit the analysis results from justifying representational claims, but claims about the information contained in the data are within reach. The first of these claims is a claim about the phenomena that the data were produced to evaluate, while the second is a claim about a regularity in the data — that is the information about task labels available from measures of BOLD signal activity. The

---

[15] This has been referred to as a 'reverse inference', a practice that has only come to be viewed as possible with the development of MVPA techniques (compare Poldrack 2006; 2011).

targets of these claims map roughly onto what Uljana Feest calls ‘hidden’ and ‘surface’ phenomena (2011).

According to Feest, surface phenomena are “... equated with empirical data patterns that are either found in the world or created in the lab…” and hidden phenomena are “... more removed from particular regularities…” (p. 63). Surface phenomena occur as data patterns, while hidden phenomena are indicated by, and so are more removed from, the data. This distinction tracks the difference between claims about the discriminability of task conditions, which are justified by patterns in the data, and claims about representations, which are distant from the data and are not justified by the results of any particular data analysis procedure. Multiple analysis procedures, or multiple data patterns, are the required evidence for claims about hidden phenomena. If not evidence that justifies an inference to claims about hidden phenomena, then what contribution do techniques like PCA make to assessments of data’s evidential value with respect to those claims?

An inference from data to a claim is a judgement of the evidential value of the data, but not all assessments of evidential value conclude with an inference. For instance, Sabina Leonelli, in a discussion of the processes involved in accessing and using data from a shared repository, notices that scientists first judge the relevance of the data, then the strength of evidence it provides for or against the hypotheses they are interested in (2009). More generally, Julian Reiss’ pragmatic account of evidence recognizes these judgements as reflecting different evidential relations that can hold between a body of evidence and a claim (2015).

Reiss argues for a theory of evidence[16] that provides “... criteria and guidelines that translate between knowledge of the facts relevant to a hypothesis and judgements about

[16] The account of evidence briefly discussed here is intended for evaluating evidence in situations that are not ‘epistemically ideal’. This is certainly the case in neuroimaging

the hypothesis" (p. 343). He refers to these as 'supporting evidence' and 'warranting evidence' respectively. Supporting evidence is evidence that a fact, data set or source of information is relevant for evaluating a hypothesis or claim. Supporting evidence can be directly or indirectly related to a claim. A result directly supports a hypothesis when the result is predicted by the hypothesis, and indirectly supports it when the result is incompatible with one or more alternative hypotheses (p. 347). Warranting evidence, or warrant, is grounds for inferring that a given hypothesis or claim is true (p. 342-3). Applying this distinction to the results of PCA diffuses the apparent tension in its use.

PCA aids the assessment of the evidential value of the data with respect to representational claims by providing supporting evidence for those claims. The support is provided via the claim about the data that is warranted by the PCA results. The kind of support provided depends on the details of the experiment. Direct support requires that the evidence, the claim about the data in this case, is predicted by the claim about the target phenomenon. Indirect support requires the claim about the data to be consistent with the target claim and inconsistent with a number of alternative claims about the phenomenon. I briefly consider examples of each.

On the minimal assumption that at least some differences in representations in the brain are "… represented by different patterns of neural firing…" (Norman et al 2006, p. 425), the claim that 'region y represents cognitive dimension x' predicts that differences in brain activity in region y can be used to discriminate between cognitive states that vary along dimension x. For example, it has been proposed that pattern classifier accuracy could be used to identify candidates for neural correlates of consciousness (Sandberg et al 2014). The logic behind this is that neural activity representing the content of an experience will be more consistently correlated with that experience, and so the BOLD activity correlated with that activity should permit more accurate and consistent

---

research where causal factors of interest cannot be intervened upon, and the data points measured are indirectly related to the phenomena of interest.

classification of a participants' awareness of stimuli (p. 4). In such a case, high classifier accuracy would provide direct support for a claim about neural correlates of consciousness, as the target claim about phenomena predicts a claim about data. That is, classification accuracy will be high for regions where candidate neural correlates occur.

The first use of pattern classification techniques in neuroimaging research provides an example of indirect support (Haxby et al 2001). In that paper, Haxby and colleagues used PCA to argue that object representations in the hippocampus are distributed throughout the region and not localized to specific areas such as the 'parahippocampal place area' or 'fusiform face area'. They use PCA to show that the brain activity could be used to discriminate the object categories of stimuli with high accuracy even when the data from the specialized sub-regions is excluded from the analysis. This result is consistent with the distributed representation hypothesis and inconsistent with the localized representation hypothesis. In this way the classifier's performance provides indirect support for the distributed processing claim.

By providing supporting evidence, directly or indirectly, for representational claims pattern classification analysis is able to assist in the assessment of the evidential value of neuroimaging data with respect to those claims. This is consistent with arguments that PCA results are unable to justifiably warrant such claims. While this accounts for the contribution that data analysis techniques make to the interpretation of neuroimaging data, it is not sufficient to explain how such inferences could justifiably be made. Fortunately, multiple data analysis techniques are used to assess the evidential value of neuroimaging data (e.g., Wright forthcoming). Each individual technique at best provides supporting evidence for claims about the target (hidden) phenomena, and only justifies claims about the data (or surface phenomena). For the inferences that the data interpretation process results in to be justified, the process must be such that identifying and interpreting a number of data patterns 'adds up' to warrant. Doing so requires the process to be sensitive to the other line of skepticism: the underdetermination due to the indirect nature of the data and complexity of the system under investigation.

3.4 Explaining Data Patterns

The data patterns that are the result of a data analysis technique are inferentially limited. They provide warrant for claims about the data, reflecting surface phenomena, but merely support claims about the hidden phenomena that are the targets of investigation. In practice, multiple techniques and variations on those techniques are used to assess the evidential value of neuroimaging data. What remains to be shown is how these results can elevate the data set they are derived from to the status of warranting evidence. Put another way, by what process of reasoning are the claims about the data inferred from data analysis results brought to bear on the target claim about phenomena?

Warrant, on Reiss' view, is a feature of a diverse body of evidence that includes both direct and indirect support. Crucial to this process is the rejection of alternatives, which is provided by indirect support (p. 357-8). As Reiss notes, rejection of alternatives is a judgement, and supporting evidence at best recommends a decision, but cannot decisively rule out all possible alternatives (p. 354). Reiss identifies a number of pragmatic criteria that scientists rely on when such judgements come into conflict. These include the effect size, study characteristics and political, social and economic considerations (p. 355-6). While enumerating pragmatic criteria is useful for explicating the factors that contribute to a given inference that is the product of an assessment of evidential value, it leaves unspecified the details of the reasoning process by which such judgements are arrived at. This is especially true in the context of criticisms of neuroimaging research, where the underdetermination of claims about phenomena challenge the adequacy of assessments of the warrant of claims based on neuroimaging data.

For example, Christopher Mole and Colin Klein argue that inferences from neuroimaging data often mistakenly treat consistency with a claim as evidence for that claim (2010, p. 101). They argue that the mere consistency of data and a hypothesis is insufficient grounds for inferring, assenting to or believing the hypothesis in question because consistency does not entail that alternatives are ruled out, and it is only by ruling out alternatives that hypotheses are confirmed (Mole and Klein 2010, p. 100). This is in line with Reiss' pragmatic account of evidence, which puts the weight of warrant on how much direct support there is for the alternatives ruled out by the indirectly supporting

evidence (2015, p. 359). Unfortunately, the indirect nature of neuroimaging data and the causal complexity of neural and cognitive systems ensure that viable alternatives appear easy to come by.

Mole and Klein argue that aspects of neuroimaging data provide ample space for constructing viable alternatives to an inferred claim. These include appealing to details of the experiment not explicitly considered in the interpretation (2010, p. 105), varying the parameters of analysis procedures (p. 108), and adopting alternative background theories (p. 109). Even if the details and conclusions of this argument are not aligned with the practice[17], they raise an important challenge for inferences in neuroimaging research. Pragmatic criteria of the form Reiss proposes will not be sufficient for addressing this challenge, as they do not address what appears to be a persistent underdetermination of the inferences. A comparison with the severity requirement in Mayo's error-statistical account of evidence is suggestive of what is missing from the story so far (1991; 1996; Mayo and Spanos 2010).

The severity requirement states that, where T is a statistical test, e is an experimental outcome or data set, and H is a hypothesis or hypothetical claim, "[p]assing a test T (with e) counts as a good test of or good evidence for H just to the extent that H fits e and T is a severe test of H" (1996, p. 180). In other words, the evidential value of data with respect to a claim is a function of the severity of the relationship between the data, a claim and the testing procedure that connects them. Severity is not an all-or-nothing feature of a test. Instead, it can be improved by eliminating alternatives, testing assumptions of the statistical models, and identifying sources of experimental error — all of which are

---

[17] Mole and Klein's argument is specifically based on the logic of null-hypothesis statistics testing, which has been shown to be more sophisticated than their treatment of it (Machery 2014), and may also make the error of treating data patterns as evidence for claims about phenomena, which, if it happens in experimental practice, is rare (Roskies 2010; Wright forthcoming).

procedures that can be conducted independently of the primary test whose severity is being evaluated (also see Mayo 1991; Mayo and Spanos 2011).

Like Reiss' account, the severity principle recognizes the importance of ruling out alternatives and accumulating evidence. Unlike Reiss' account, it situates the 'test', or more generally the reasoning process by which the evidential relation is assessed, as a necessary component of the criteria for evaluating the resulting inferences. What's missing, then, is an account of the reasoning process neuroscientists engage in when they assess the degree of warrant a neuroimaging data set provides for a claim about the phenomenon it was produced to investigate. To aid in articulating that process I briefly consider an example.

3.4.1 Representations in Parahippocampal Cortex

Martin and colleagues used pattern classification analysis to bring neuroimaging data to bear on claims about the content of memory representations in subregions of the hippocampus. At the time this paper was published, the received view was that PhC's contribution to memory was representing episodic contextual information. Contrary to this view, these authors argue that "… parahippocampal (PhC) cortex does not only represent episodic context but can also represent item information for some object categories in recognition-memory decisions" (Martin et al 2013, p. 10915).[18]

The particular memory phenomenon they are interested in is familiarity, which is often distinguished from recollection. A recollection is a memory of an object that includes contextual details of the encounter, while familiarity is a memory of an object absent such contextual details. That is, a 'feeling' that an object is familiar, without explicit memory of what else occurred in the previous encounter with the object that is being remembered.

---

[18] The study also examines representations in other regions in the medial temporal lobe (MTL), but for the sake of space I focus on only one claim.

To investigate memory representations in PhC, they use a two stage experimental task. In the first stage, participants are presented with images of faces, chairs and buildings, and have to decide for each if it was attractive, comfortable or valuable respectively (p. 10916). This stage is conducted outside of the scanner. In the second stage, participants are placed in an MRI scanner and functional data are collected while they perform a recognition memory task. They are presented with images from each category (face, chair and building), some of which they had seen in the first stage, and some of which are novel. For each image, they rate their familiarity with the item on a scale of 1 (least familiar) to 4 (most familiar). A separate response option allows participants to indicate if they recollect contextual details associated with the feeling of familiarity so that these data points can be excluded from the analysis.

There are two aspects to the claim that the regional activity contains category-specific memory representations. That is, the representations are both category-specific and influence memory. To assess if the data so produced warrants the target claim, they use pattern classification analysis to address two central questions. The first is to "… determine whether distributed patterns of activity in any of the [medial temporal lobe] structures examined could reliably distinguish between the stimulus categories" and the second is to "… examine whether distributed patterns of activity could be identified that reflected a memory signal, ie. differences between familiar and novel stimuli, for each stimulus category" (p. 10917). The first question addresses the category-specificity, and the second the influence on memory. Notice that neither question is a question about the phenomena per se, but questions that pertain to patterns in the data. These questions are informed by consideration of the phenomena under investigation, the design of the experimental task, and are formulated as questions that can be addressed using a pattern classification method.

To address the first question, a classifier is trained to label the stimulus category of the objects based on the functional data. To ensure that the underlying signal could not be associated with memory the analysis includes only data from the novel trials. The resulting pattern shows that the classifier performs above chance at labelling the object category using data from the regions of interest (p. 10918). This result is reinforced by

pair-wise classification analysis for each region and each pairing of the categories (buildings vs chairs; buildings vs faces; faces vs chairs). These patterns show classifier accuracy is above chance at discriminating between any pair of categories in PhC (p. 10919). This together is taken to be "… evidence for category specific representations…" in PhC. The patterns isolated by these analyses are restricted to novel trials, and so they do not assist in determining the evidential status of the data with respect to the memory aspect of the central claim.

To address the second question, the classifier is trained to label each trial as low (1 or 2) or high familiarity (3 or 4). Classification is then performed for each object category. The resulting patterns show that activity in PhC could be classified by familiarity rating only for buildings and chairs, but not for faces (p. 10919). These results, together, support the claim that PhC represents category information for buildings and chairs in the context of memory judgements.

Examining this set of three patterns suggests an alternative explanation for the result. Specifically, "… above-chance classifier performance … [could be] based on a common familiarity signal" (p. 10920). That is, while there is evidence for category relevant information, the results up to this stage provide no evidence that the information is actually category-specific. It could be that the familiarity signal is shared across categories, it is not 'chair' or 'building' specific but instead reflects a category-independent feature of the stimuli in a manner that the classifier can leverage to label the data. To address this possibility, they use a cross-classification analysis. This involves training the classifier to label the data as familiar or novel within one category, then testing it on data for a different category. If the information permits category discrimination but is not category-specific, then the classifier should perform above chance in cross-classification. Just as the decision to use PCA in the first place is guided by the target phenomena, here too the decision to perform a cross-classification analysis is guided by the search for a pattern that would be explained by only one of the candidate claims. In the end, the cross-classification test result supports the claim that the information is category-specific, as when the classifier is trained on building data it fails to reliably label chair data and vice versa (p. 10920).

These patterns (and others I have not discussed) are taken to justify the judgement that the data are evidence for the claim that PhC represents object-category information in the context of memory judgements. Notice that it's not a single classifier result, but a collection of them, that provides warranting evidence for the claim. The process of incorporating diverse analysis results into a broader interpretation of the data is driven by attempts to explain the analysis results in terms of claims about phenomena. This includes a search for patterns that provide evidence that the data are relevant for addressing the question, further patterns to work towards developing a clear answer to questions relating to the target claim, and patterns specifically for the purpose of ruling out alternative explanations for the set of patterns isolated up to that stage. Treating the reasoning process involved in assessments of neuroimaging data as explanatory provides the resources necessary to weaken the threat of underdetermination.

3.4.2 Towards the Best Explanation

I. J. Good's discussion of the reasoning process involved in exploratory data analysis is, perhaps surprisingly, a good fit for capturing how data analysis results contribute to assessments of the evidential value of neuroimaging data. Good argues that exploratory data analysis involves manipulating the data to highlight specific features and suppress others (1986, p. 290). The aim of exploratory analysis is to identify patterns in the data that are potentially explicable, then to formulate hypotheses about those patterns and improve those by examining other patterns in the data (p. 291). When a plausible explanation for a pattern is identified, the data analyst examines the residuals of the pattern to evaluate the plausibility of the offered explanation. The residuals are the portions of data that were set aside, or suppressed, by the manipulations used to arrive at the original pattern. The same is true of the interpretation of neuroimaging data.

The investigators in the case discussed above chose to use PCA because of its potential to isolate patterns in the data that could be explained by the target phenomenon. The patterns are regarded as potentially explicable in terms of claims about representations in part because the form of patterns that obtain indicates which of the claims under consideration are likely to be true. Once a number of patterns are isolated, explaining them leads investigators to notice alternative explanations not controlled for in the data

production process. The potential availability of these alternatives threatens the plausibility of the offered explanation, and so they isolate and examine other patterns in the data to rule out alternatives. All of this was done while actively considering the experimental protocols involved in the production of the data, the manipulations implemented by the classifier, and the overall aim of the research project.

The judgement that neuroimaging data warrants a claim about a phenomenon is justified when the claim can be shown to be the best (available) explanation for the data. This is established by isolating patterns in the data that are explicable in terms of the claims under consideration. Warranted claims are those that explain the collection of isolated data patterns, while also taking into account the procedures of data production and manipulations that together produced each data pattern. Claims about data provide useful scaffolds for this reasoning process, as claims about phenomena can sometimes be used to formulate claims about data in the form of a prediction, and because claims about data are, unlike the abstract data patterns that justify them, more susceptible to explanations.

An inference in neuroimaging research is warranted not only by the structure of the body of evidence that the accumulated data patterns provide, but also to the degree that the claim inferred is regarded as the best explanation of the data patterns isolated in the process of interpretation. This provides some resistance against straightforward underdetermination arguments such as that offered by Mole and Klein. There are, as Mole and Klein notice, a variety of decisions made in the process of producing and analyzing neuroimaging data, and any given decision may be the cause of an observed data pattern. However, data patterns are not considered in isolation of the rest of the experimental context, and so alternative hypotheses based on isolating one or another decision point as critical and mistaken are not likely to be serious threats to the inferences scientists make. Alternatives, to be viable threats to an inference, must be as good of an explanation of the total set of patterns explained by the claim under scrutiny. Furthermore, where alternatives are threats, as in the case considered above with respect to the possibility that the familiarity signal is shared across categories, it is sometimes possible to rule them out, or at least address them, by analyzing the data further. One way this is done is by articulating the data patterns predicted by the viable alternative that can

be isolated with the available tools of data analysis, and conducting the relevant analyses. If those patterns are not found, the once-viable alternative can no longer be regarded as sufficient to explain the collected data patterns.

Explanations are valuable because they are easier to communicate and conceptualize in terms of claims about phenomena than statistical facts, or raw data. Their function is heuristic, and not necessarily truth conducive. In this regard, I am sympathetic to Andrea Woody's recent work on explanation in science, in which she argues that explanations are "… genuinely important to the proper functioning of science, but it is a worker bee, rather than … a shiny bauble proudly displayed in the aftermath of scientific activity, or … a mysterious seer pointing us toward a yet undiscovered truth" (2015, p. 86). On her view, philosophical inquiry into explanations ought to begin by asking what roles they play in science. Here, I have argued that they are used in making judgements about the evidential value of data.

3.5 Conclusion

Treating the process of data interpretation as explanatory allows the well-known challenges associated with using neuroimaging data to be reconciled with the continued enthusiasm for their potential. Data analysis techniques are used to isolate patterns in data that are explicable in terms of the target claim. Insofar as these claims are consistent with claims about phenomena they are evidence that the data set they are about is relevant to consider in evaluations of those claims. As more patterns are isolated, investigators attempt to explain the collection of patterns in terms of claims about the target phenomena. Neuroimaging data are judged to warrant a claim about the relationship between brain activity and cognition to the extent that such a claim explains the collection of data patterns isolated in the process of its interpretation. This process weakens threats of underdetermination due to the indirectness of data and complexity of the system because a viable alternative cannot simply explain a single pattern, but must account for at least as many of the patterns as the claims inferred from the data explain.

If the purpose of isolating a data pattern is to arrive at a form of evidence that is interpretable by a human investigator, as Good argues, then an important contribution to

the epistemology of neuroimaging research would be to examine how the process of data analysis can lead well-intentioned investigators to misinterpret the resulting pattern. Treating this process as explanatory suggests that what is required is an account of what criteria neuroscientists do (and ought) to use in evaluating competing explanations of data patterns. This is particularly important given current trends in neuroimaging research.

New data analysis techniques are continually in development and, especially with the rising popularity of multivariate methods, each is more complex and sophisticated than the last. While new analysis techniques certainly constitute a form of progress, this progress is regarded with caution by many members of the neuroscientific community. Concerns raised include a trade off in computational sophistication and deeper understanding of the material objects under investigation, that is brains (Marder 2015), and inferential risks that are associated with an increase in analytic flexibility within a scientific community which can lead to an increase in false positive rates (Carp 2012). Proposals for improving the epistemic status of neuroimaging research promote sharing data and sharing algorithms (Poldrack et al 2017), and include a demand for pre-registration of research plans (Munafò et al 2017). The explanatory treatment of data interpretation argued for above suggests that the success of these models and research strategies will depend on how they interact with the criteria communities, and individual researchers, use to assess the adequacy, or 'bestness', of explanations of data patterns.

Engaging with concerns about analytic flexibility, the differential expertise with respect to the computational and material aspects of neuroscience, and the promise and challenges with reproducible neuroscience are important and valuable areas for philosophical inquiry. As these issues are timely, philosophical contributions could have a direct impact on the trajectory of cognitive neuroscience. These impacts, however, are only possible to the extent that philosophical analyses are sensitive to the relevant details of the practice, and how the procedures that compose that practice contribute to its outcomes. I offer the account provided here as one way to approach such an analysis.

References

Aktunç, M. E. [2014a]: 'Severe tests in neuroimaging: what we can learn and how we can learn it', Philosophy of Science, 81(5), pp. 961-973.

Aktunç, M. E. [2014b]: 'Tackling Duhmeian Problems: An Alternative to Skepticism of Neuroimaging in Philosophy of Cognitive Science', Review of Philosophy and Psychology, 5(4), pp. 449-464.

Anderson, M. L. and Oates, T. [2010]: 'A critique of multi-voxel pattern analysis', Proceedings of the 32nd Annual Conference of the Cognitive Science Society, pp. 1511-16.

Ashby, F. G. [2011]: Statistical Analysis of fMRI Data. The MIT Press.

Bogen, J. [2001]: 'Functional imaging evidence: Some epistemic hotspots' In Peter K. Machamer, Peter McLaughlin & Rick Grush (eds.), Theory and Method in the Neurosciences. University of Pittsburgh Press. pp. 173-199.

Bogen, J. [2002]: 'Epistemological custard pies from functional brain imaging', Philosophy of Science, 69, pp. S59-S71.

Bogen, J. and Woodward J. [1988]: 'Saving the Phenomena', Philosophical Review, 97, pp. 303-52.

Carp, J. [2012]: 'On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments', Frontiers in Neuroscience, 6, pp. 149.

Davis, T., LaRocque, K. F., Mumford, J. A., Norman, K. A., Wagner, A. D., and Poldrack, R. A. [2014]: 'What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis', Neuroimage, 97, pp. 271-83.

Davis, T. and Poldrack, R. A. [2013]: 'Measuring neural representations with fMRI: practices and pitfalls', Annals of the New York Academy of Sciences, 1296, pp. 108-34.

de-Wit, L. Alexander, D., Ekroll, V., and Wagemans, J. [2015]: 'Is neuroimaging measuring information in the brain?', Psychonomic Bulletin and Review, 23, pp. 1415-1428.

Etzel, J. A., Zacks, J. M., and Braver, T. S. [2013]: 'Searchlight analysis: promise, pitfalls, and potential', Neuroimage, 78, pp. 261-269.

Feest, U. [2011] 'What exactly is stabilized when phenomena are stabilized?', Synthese, 182, pp. 57-71.

Good, I. J. [1983]: 'The philosophy of exploratory data analysis', Philosophy of science, 50(2), pp. 283-295.

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. [2003]: 'Functioanl connectivity in the resting brain: a network analysis of the default mode hypothesis', Proceedings of Natural Academy of Science USA, 7, pp. 253-258.

Hacking, I. [1983]: Representing and Intervening: Introductory topics in the philosophy of natural science, Cambridge University Press.

Hardcastle, V. G., & Stewart, C. M. [2002]: 'What do brain data really show?', Philosophy of Science, 69(S3), pp. S72-S82.

Havlicek, M., Roebroeck, A., Friston, K., Gardumi, A., Ivanov, D., and Uludag, K. [2015]: 'Physiologically informed dynamic causal modeling of fMRI data', NeuroImage, 122, pp. 355-372.

Haxby, J.V., Connolly A.C., and Guntupalli J.S. [2014]: 'Decoding Neural Representational Spaces Using Multivariate Pattern Analysis], Annual Review of Neuroscience, 37.

Haxby, J. V. [2010]: 'Multivariate Pattern Analysis of fMRI data' in M. Bunzl and S. J. Hanson (eds), Foundational Issues in Human Brain Mapping, The MIT Press, pp. 55-68.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai A., Scouten, J. L., and Pietrini, P. [2001]: 'Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex', Science, 293, pp. 2425-30.

Haynes, J. [2012]: 'Brain reading', in Richmond, S. D., Rees, G., and Edwards, S. J. L. (eds.), I Know What You're Thinking: Brain imaging and mental privacy, Oxford University Press.

Kamitani, Y., and Tong, F., [2005]: 'Decoding the visual and subjective contents of the human brain', Nature Neuroscience, 8, pp. 679-685.

Kim, G., Lewis-Peacock, J. A., Norman, K. A., and Turk-Browne, N. B. [2014]: 'Pruning of memories by context-based prediction error', Proceedings of the National Academy of Sciences, 111(24), pp. 8997-9002.

Kohler, P.J., Fogelson, S.V., Reavis, E.A., Meng, M., Guntupalli, J.S., Hanke, M., Halchenko, Y.O., Connolly, A.C., Haxby, J.V. and Peter, U.T., [2013]: 'Pattern classification precedes region-average hemodynamic response in early visual cortex'. NeuroImage, 78, pp. 249-260.

Klein, C. [2012]: 'Cognitive Ontology and Region- versus Network-Oriented Analyses', Philosophy of Science, 79(5), pp. 952-960.

Klein, C. [2010]: 'Images are not the evidence in Neuroimaging', British Journal for the Philosophy of Science, 61, pp. 265-78.

Kriegseskorte, N., Mur, M., and Bendettini, P., [2008]: 'Representational similarity analysis - connecting the branches of systems neuroscience', Frontiers in Systems Neuroscience, 24.

Landreth, A. and Richardson, R. C. [2004]: 'Localization and the new phrenology: A review essay on William Uttal's the new phrenology', Philosophical Psychology, 17, pp. 107-23.

Leonelli, S. [2015]: 'What counts as scientific data? A relational framework', Philosophy of Science, 82(5), pp. 810-821.

Leonelli, S. [2009]: 'On the locality of data and claims about phenomena', Philosophy of Science, 76(5), pp. 737-749.

Liu, T., Hospadaruk, L., Zhu, D. C., and Gardner, J. L. [2011]: 'Feature-Specific Attentional Priority Signals in Human Cortex', The Journal of Neuroscience, 31, pp. 4484-95.

Logothetis, N. K. [2008]: 'What we can do and what we cannot do with fMRI', Nature, 453(7197), pp. 869-878.

Machery, E. [2014]: 'Significance Testing in Neuroimagery', in Kallestrup J., and Sprevak, M.(eds.), New Waves in the Philosophy of Mind, Palgrave Macmillan, pp. 262-277.

Marder, E. [2015]: 'Understanding Brains: Details, Intuitions, and Big Data', PLoS Biology, 13(5), e1002147.

Martin, C. B., Cowell, R. A., Gribble, P. L., Wright, J., Köhler, S. [2015]: 'Distributed category-specific recognition memory signals in human perirhinal cortex', Hippocampus.

Martin, C. B., McLean, D. A., O'Neil, E. B., and Köhler S. [2013]: 'Distinct Familiarity-Based Response Patterns for Faces and Buildings in Perirhinal and Parahippocampal Cortex', The Journal of Neuroscience, 33, pp. 10915-23.

Mayo, D. [1996]: Error and the Growth of Experimental Knowledge, The University of Chicago Press.

Mayo, D. [1991]: 'Novel Evidence and Severe Tests', Philosophy of Science, 58(4), pp. 523-552.

Mayo, D., and Spanos, A. [2010]: 'Error Statistics', in Bandyopadhyay, P. S., and Forster, M. R. (eds.), Handbook of the Philosophy of Science. Vol 7: The Philosophy of Statistics. Elsevier, pp. 153-198.

Misaki, M., Kim, Y., Bandettini, P. A., and Kriegeskorte, N. [2010]: 'Comparison of multivariate classifiers and response normalizations for pattern- information fMRI', NeuroImage, 53, pp. 103-18.

Mole, C. and Klein, C. [2010]: 'Confirmation, Refutation, and the Evidence of fMRI', In Stephen Hanson & Martin Bunzl (eds.), Foundational Issues in Human Brain Mapping. Cambridge: MIT Press. pp. 99-112.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E., Ware, J. J., and Ioannidis, J. P. A. [2017]: 'A manifesto for reproducible science', Nature Human Behaviour, 1.

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby J. V. [2006]: 'Beyond mind-reading: multi-voxel pattern analysis of fMRI data', Trends in Cognitive Sciences, 10(9), pp. 424-430.

Pereira, F., and Botvinick M. [2011]: 'Information mapping with pattern classifiers: a comparative study', Neuroimage, 56(2), pp. 476-496.

Pereira, F., Mitchell, T., and Botvinick M. [2009]: 'Machine learning classifiers and fMRI: A tutorial overview', Neuroimage, 45, pp. 199-209.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò M. R., Nichols, T. E., Poline, J., Vul, E., and Yarkoni, T. [2017]: 'Scanning the horizon: towards transparent and reproducible neuroimaging research', Nature Reviews Neuroscience, 18, pp. 115-126.

Poldrack, R. A. [2011]: 'Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding', Neuron, 72(5), pp. 692-697.

Poldrack, R. A. [2006]: 'Can cognitive processes be inferred from neuroimaging data?', Trends in Cognitive Sciences, 10(2), pp. 59-63.

Poldrack, R. A., Mumford J. A., and Nichols, T. E. [2011] Handbook of Functional MRI Analysis, Cambridge University Press.

Radder, H. ed. [2003]: The philosophy of scientific experimentation. University of Pittsburgh Press.

Reiss, J. [2015]: 'A pragmatist theory of evidence', Philosophy of Science, 82(3), pp. 341-362.

Rheinberger, H. [2011]: 'Infra-Experimentality: From Traces to Data, From Data to Patterning Facts', History of Science, 49(3), pp. 337-348.

Ritchie, J.B., Kaplan, D.M., and Klein, C. [Forthcoming]: 'Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience', British Journal for the Philosophy of Science.

Roskies, A. [2010]: 'Neuroimaging and Inferential Distance: The Perils of Pictures', in M. Bunzl and S. J. Hanson (eds), Foundational Issues in Human Brain Mapping, The MIT Press, pp. 195-216.

Sandberg, K., Andersen, L. M., and Overgaard, M. [2014]: 'Using multivariate decoding to go beyond contrastive analysis in consciousness research', Frontiers in Psychology, 5, pp. 8-13.

Seiger, R., Hahn, A., Hummer, A., Kranz, G.S., Ganger, S., Küblböck, M., Kraus, C., Sladky, R., Kasper, S., Windischberger, C. and Lanzenberger, R., [2015]: 'Voxel-based morphometry at ultra-high fields. A comparison of 7T and 3T MRI data', NeuroImage, 113, pp. 207-216.

Sullivan, J. [2009]: 'The multiplicity of experimental protocols: A challenge to reductionist and non-reductionist models of the unity of neuroscience', Synthese, 167, pp. 511-39.

Tong, F., and Pratte, M. S. [2012]: 'Decoding Patterns of Human Brain Activity', Annual Review of Psychology, 63, pp. 483-509.

Uttal, W. [2001]: The New Phrenology, The MIT Press.

van Orden, G. C., and Paap, K. R. [1997]: 'Functional Neuroimages Fail to Discover Pieces of Mind in Parts of the Brain', Philosophy of Science, 64, pp. S85-94.

Woodward J. [2000]: 'Data, phenomena, and reliability', Philosophy of Science, 67(3), pp. S163-S179.

Woody, A. [2015]: ‘Re-oriented discussions of scientific explanation: A functional perspective”, Studies in History and Philosophy of Science, 52, pp. 79-87.

Wright, J. W. [Forthcoming]: ‘The Analysis of Data and the Evidential Scope of Neuroimaging Results’, British Journal for the Philosophy of Science.

Chapter 4

4 Meta-Analyses and Brain Mapping

4.1 Introduction

Cognitive neuroscience, broadly understood as the study of the relationship between the brain and cognition, has seen a significant increase in the volume and complexity of data produced over the last two decades. The rapid growth of subfields and steady increase in the sophistication and power of tools for the production and analysis of data has added to existing concerns about the effectiveness of the knowledge produced in cognitive neuroscience. The high volume of published results makes it difficult for investigators to stay aware of current trends even within their own subfield (Yarkoni et al 2011; Silva, Bickle and Landreth 2013). Indeed, the isolation of areas of research, which are organized around investigations of particular regions of the brain, or particular cognitive functions, has been flagged as a possible cause for the lack of critical discussions about the adequacy of popular brain mapping strategies (e.g., Price and Friston 2005). Additionally, neuroscientists are increasingly aware of the limitations of neuroimaging experiments, which, due to practical constraints, have low statistical power and are unable to support generalized claims about the relationship between brain activity and cognitive functioning (Lieberman and Cunningham 2009; Costafreda 2011; Poldrack et al 2017). In light of these challenges, philosophers and neuroscientists have identified large-scale meta-analyses as a promising method for making progress in cognitive neuroscience (e.g., Yarkoni et al 2011; Klein 2012; Silva, Bickle & Landreth 2013; Anderson 2015; Sullivan 2017).

Meta-analyses are already being used to address a number of barriers impeding progress in neuroimaging research. These challenges range from the low statistical power of inferences, to concerns about the validity of concepts used to theorize about cognitive processes and states (Yarkoni et al 2011; Anderson 2015). The effectiveness of a meta-analysis is partly tied to the volume and diversity of data that it operates over. Thus, the

push for bigger, more powerful and more robust meta-analyses is one of the primary motivations behind the development and maintenance of repositories of neuroimaging data, such as NeuroSynth (Yarkoni et al 2011), BrainMap (Laird 2011), NeuroVault (Gorgolewski et al 2015), openfMRI (Poldrack et al 2013) and even a database of meta-analysis results, ANIMA (Reid et al 2015).

Philosophical work on the epistemic impacts of and conditions necessary for data sharing and data repositories to promote knowledge generation shows the importance of community coordination and collaboration (e.g., Leonelli and Ankeny 2012; O'Malley and Soyer 2012; Leonelli 2016). Recent work on the prospects of meta-analyses and large repositories of neuroimaging data indicates that cognitive neuroscience presently lacks sufficient coordination of the methodological and conceptual practices necessary to promote knowledge generation through data sharing (Sullivan 2017). Complementing concerns about the need for community coordination and organization, here I argue that research practices themselves must change with the uptake of meta-analysis tools and widespread use of shared data sets.

If not accompanied by changes in the practices and standards that inform and direct the analysis and interpretation of data, easy access to large-scale meta-analyses could lead to the production of apparently justified, but likely misleading theories and hypotheses. To make this argument I examine a recent dispute over the evidential significance of meta-analysis results between well-intentioned users of the NeuroSynth database, a data set made by the automated extraction of reported findings from published neuroimaging studies, and the lead developer. I argue that the dispute reflects a difference in the methods of analysis and interpretation favoured by the investigators, as well as a different understanding of the implications of data manipulations used to synthesize data drawn from diverse sources. The inferential and analysis errors the database users make indicates an inadequacy in the approach to data interpretation that they take. Their approach is sound in the context of neuroimaging research, but likely to be mistaken when applied to the analysis of a data set made from the synthesis of a diverse collection of neuroimaging results. If it is to be epistemically advantageous, the development and uptake of data repositories and the large-scale meta-analyses they make possible must be

accompanied by the development and uptake of research practices adequate for the analysis and interpretation of diverse data sets.

I proceed as follows: in section 2 I begin by outlining some of the problems meta-analyses and repositories of neuroimaging data are expected to solve. In section 3 I briefly discuss challenges for meta-analyses that follow from current research practices operative in neuroscience. In doing so, I raise the possibility that changes in research practices may be necessary if meta-analysis techniques and data repositories are to succeed at resolving the problems outlined in the second section. In section 4 I review a dispute over the interpretation of meta-analysis results, emphasizing the factors that direct the use and interpretation of the shared dataset and contribute to the persistent disagreement between participants in the dispute. In section 5 I argue that research practices in cognitive neuroscience need to be adapted in order for meta-analyses and data repositories are to promote genuine progress.

4.2 Meta-Analyses, Databases and their Promise

Cognitive neuroscience, and neuroimaging research in particular, faces a number of practical, theoretical and structural barriers to progress. These problems include the inability of neuroimaging experiments to provide evidence that can support generalizable claims about the relations between brain activity and cognitive states, and the rapidly increasing volume of publications in the primary literature. Aggregating data from disparate experiments into accessible repositories, and providing meta-analysis tools to evaluate their evidential significance, are expected to address these barriers to the production of knowledge from the results of neuroimaging experiments.

Functional magnetic resonance imaging (fMRI), is used to measure the ratio of oxygenated to deoxygenated hemoglobin in small pieces of the brain (called voxels, which are typically 2 to 3 mm cubes). This is called the BOLD (blood oxygenation level dependant) signal, and it is used as an indirect measure of neural activity. A change in the BOLD signal within a voxel occurs when the oxygenation needs of cells in that voxel change. As neural activity increases, the neurons (and other cells) will require more oxygenation and so the BOLD signal will change. Neuroimaging data are collected while

a human participant performs cognitive tasks. Tasks are designed to require participants to engage specific cognitive processes, such as memory or attention, or experience specific cognitive states, such as a feeling of familiarity or pain. The resulting data set consists in BOLD signal measurements that are correlated with the performance of specific cognitive tasks.

Investigators go to great lengths to ensure that only the cognitive processes of interest are reflected in an experimentally produced data set. Such a data set may be useful for identifying the origin and character of the brain activity sufficient for the implementation of the particular cognitive states or processes of interest, but is unable to substantiate claims about the general function performed by the brain region or network. This problem is at the heart of discussions about the evidential value of neuroimaging data for reverse inferences, which involve attributing cognitive states or processes on the basis of observed patterns of brain activity (see Poldrack 2006, 2011; Klein 2012; Machery 2014; Glymour and Hanson 2016).

Neuroimaging data showing, for example, a correlation between the experience of pain and activity in the dorsal anterior cingulate cortex (dACC), cannot warrant the corresponding reverse inference claim that if 'a subject's dACC is active, then they are in pain'. The central problem with the reverse inference claim is that brain regions[19] could also be activated by a variety of different cognitive tasks and processes. Data originating from a neuroimaging experiment correlating pain and dACC activity, on their own, cannot substantiate claims about reverse inferences because such a data set does not reflect a sufficiently diverse collection of cognitive states and processes. For instance, in

---

[19] While there is hope that network-analyses may be able to resolve this problem (e.g., Klein 2012), there is evidence that network activity also stands in a many-to-many relationship with cognitive processes (see Pessoa 2014 and Horwitz 2014 for a relevant discussion). For this reason, while I focus this discussion on brain regions and areas, it is likely that the same concerns and challenges will also apply to network-based approaches.

addition to pain, experimental data have correlated dACC activity with executive control (Carter et al 1999), emotion (Etkin et al 2011), and salience (Menon and Uddin 2010). Thus, an observation of dACC activity, at best, warrants a claim that one, or some combination of, these cognitive states or processes are active.

The problems with reverse inferences are barriers to achieving a larger theoretical goal of cognitive neuroscience: mapping the relationship between brain activity and cognitive processes, states and capacities. Currently competing theories of the relationship between brain activity and cognition propose everything from one-to-one mappings in which each brain activity pattern has an associated cognitive process or state, to many-to-many mappings in which activity patterns and cognitives states have a complex relationship. These theories have similar evidential requirements that neuroimaging experiments fall short of providing. Evidence for claims about the relationship between brain activity patterns and cognitive processes must include evidence that the cognitive process of interest is associated with a brain activity pattern, called a forward inference, and that the brain activity pattern indicates the engagement of the cognitive process, or evidence for the corresponding reverse inference (Price and Friston 2005; Klein 2012; Anderson 2015). While consulting the broader literature, as briefly done above, may be the obvious solution to this problem, the steadily increasing volume of primary research findings makes that approach less feasible by the day.

Even relatively narrow subfields, such as research on ‘visual working memory’— which is a particular subcategory of working memory, which is a specific category of explicit memory — can produce over 800 publications in one year (based on a pubmed search of articles published in 2014). The same is true of research relevant to claims about the functional role of discrete brain regions (almost 8000 articles were published on the hippocampus in 2014, one of the most widely studied regions of the brain. At least 1100 of those publications include fMRI data). The volume of literature directly relevant to any given research question is only part of the problem. These searches only capture studies that explicitly report the data and findings as relevant to understanding memory or the function of the hippocampus. Countless other studies in the literature may include data patterns and correlations linking hippocampus activity and other cognitive process, and

even hippocampus activity and various memory processes, that go unreported because the hippocampus was not the region of interest in those studies.

The aggregation of data into central repositories, and the meta-analyses they make possible, are expected to provide a way through these obstructions. The hope is that, by performing meta-analyses on large collections of neuroimaging data, the statistical power of results will increase, false positives can be identified (Lieberman and Cunningham 2009; Yarkoni et al 2011; Costafreda 2011), the current status of hypotheses and theories can be established beyond the boundaries of specific research projects pursuing them (Silva, Bickle, Landreth 2013), an evidential base that can be used to evaluate generalizable claims about brain-cognition relations will become available (Price and Friston 2005; Kober and Wager 2010), and patterns in data relevant for evaluating theories beyond those the data produced to evaluate may be discovered (Van Horn and Gazzaniga 2013). Most of the advantages meta-analyses are purported to have over individual experiments rests on the diversity of the data they operate on. This is the feature that is supposed to allow neuroscientists to move away from claims restricted to the occurrences in laboratories and towards generalizable theories of the brain's contribution to the realization of cognitive capacities.

By combining data from multiple experiments each using different task manipulations, meta-analyses accommodate a diverse range of cognitive functions (Reid et al 2015; Poldrack and Gorgolewski 2015). This makes it possible to evaluate if a pattern of brain activity, or the activation of a specific area of the brain, is consistently associated with a given cognitive process, and to determine if it is specific to one cognitive process or if it is also associated with other cognitive processes. To do this is to assess the consistency and specificity of the relationship between a pattern of brain activity and a cognitive process or state. That is, to evaluate the evidential support for both forward and reverse inference claims (Kober and Wager 2010). Providing a data set and analysis tools that can support claims about forward and reverse inferences is one of the aims of the NeuroSynth's database (see Yarkoni et al 2011).

NeuroSynth data are curated by an algorithm that scans articles published in journals it can parse and extracts peak-activation coordinates from tables in those articles. Peak activation coordinates are the X-Y-Z coordinates in a brain atlas at which investigators measured the greatest level of activation in their study. A text-analysis of the abstract is used to label the coordinates. All terms (a term is a word or word pair) in the abstract that occur in a sufficient number of other abstracts in the database are assigned as labels (examples include 'pain', 'language', 'magnetic', 'task', 'working memory' and 'cingulate'). The result is a collection of peak activation coordinates labelled by words that appear in the abstract of the article the coordinates were extracted from. NeuroSynth also includes a number of automated meta-analysis tools, making it relatively easy to conduct meta-analyses of the represented literature.

These tools provide forward and reverse inference maps correlating terms and activation coordinates. A forward inference map highlights coordinates that are more likely to be reported in articles that are labelled by a given term than in articles that are not labelled by the term. For example, if you were to randomly select a 'pain' study, the coordinates indicated on the forward inference map for 'pain' are likely to be amongst the peak activation coordinates reported in the study. A reverse inference map highlights terms that are more likely to be reported in articles reporting the given co-ordinates as active than in articles without. If you were to select a random study that reported activity in a coordinate shown in the reverse inference map, it is likely that the study would have the label you ran the analysis on.

Meta-analyses and data repositories together can help address the practical challenge of canvassing an ever-growing literature by providing tools that can expedite and guide the literature review process. This is one of the functions of NeuroSynth, which is expected to "... accelerate progress in cognitive neuroscience through greater formal synthesis of the rapidly growing primary literature" (Yarkoni et al 2011, p. 489). NeuroSynth not only aggregates data, but allows users to group and identify published research by similarities in terminology, or by similarities in reported peak activity co-ordinates. More sophisticated meta-analytic tools for collecting research outputs and providing automated guidance on future research are on the horizon. One example is the Network of

Experiments (NEX) framework presented by Silva, Bickle and Landreth (2013). The NEX uses graph-theoretic representations of neuroscientific research to evaluate the strength of evidence supporting claims about causal relations between cognitive phenomena. While there are reasons to be skeptical that the NEX framework will catalyze a revolution in neuroscience in the way the authors imagine (see Klein 2014), it is another example of the conviction that meta-analytic tools will propel neuroscience (and neuroimaging) forward by enabling the automated synthesis, search and classification of evidence available in the literature.

Whether or not these tools can succeed at achieving these goals depends on how they are received and used by the neuroscience community. In the next section I briefly review a number of challenges for the efficacy of meta-analyses that follow from concerns about the current state of experimental practice in neuroscience. I then raise it as a possibility that some of the conditions that have established some of the problems that they are supposed to resolve, are also barriers to the capacity of meta-analyses to resolve those very problems. In the following sections, by analyzing the factors contributing to a dispute over the evidential significance of NeuroSynth data, I argue that research practices must be adapted to the differences between synthesized data sets and the experimentally produced data they are derived from for meta-analyses to succeed at improving the knowledge producing capacity of cognitive neuroscience.

4.3 Challenges for Meta-Analyses

The cultivation of a diverse data set, made by combining data produced in different experiments each aiming to better understand distinct cognitive phenomena, is what enables meta-analyses to provide support for hypotheses about the relationship between cognitive processes and brain activity that are generalizable beyond the few cognitive states represented in a single, experimentally produced, data set. However, the differences between data sets that are the source of strength for meta-analyses may also be a potential weakness if research practices, and specifically approaches to the analysis and interpretation of data, are not adapted to account for the limitations and features of the data sets that result from the synthesis of disparate data.

Research programs in cognitive neuroscience are typically aimed at providing evidence for theories about the role played by regions, or networks, of the brain and specific cognitive capacities. Different investigators working in different laboratories are likely to use different task parameters, and even different tasks, to study what they consider to be the same phenomenon (i.e., 'working memory', 'cognitive control', or 'pain'). Inconsistencies due to variations in research methods and the use of terminology, especially across subdomains, are recognized as a barrier to the effective integration of data. For instance, Poldrack and colleagues note that 'working memory' has at least three distinct meanings in the literature (Poldrack et al 2011). This makes it challenging to combine data together with the aim of, for instance, conducting a meta-analysis on research related to working memory. The situation is complicated by the common treatment of tasks as equivalent to the cognitive constructs that they are used to study. Tasks typically require participants to perform a number of different cognitive processes to complete them, including the process investigators are interested in. Treating a task as equivalent to a psychological construct invokes assumptions about the cognitive strategies participants use to complete the task. The assumptions weaken the inferences drawn on the basis of the data, as they are rarely justified. The inconsistent use of terminology and treatment of tasks and constructs as equivalent makes it "… difficult to draw meaningful inferences from existing literature and limits the cumulative value of the knowledge represented in this literature" (Poldrack et al 2011). Furthermore, these problems are a recognized challenge for the integration of data in the neurosciences (Turner and Laird 2012; Sullivan 2017, p. 1-3).

The terminological ambiguity noted above is not just due to different communities of researchers using similar terms for different purposes. It is also due to the widespread disagreement over the best explanations for cognitive phenomena, and the conceptual resources necessary for understanding these phenomena. This is reflected in the structure of the Cognitive Atlas, a wiki-inspired knowledge base, which, it is hoped, will provide a framework for clarifying current descriptions of phenomena. One valuable feature of the Atlas is its wiki-like structure which allows for disagreements about the meaning and empirical basis for terms to be captured and discussed. This is important in a field where

"there is precious little consensus … regarding the basic units of mental function" (Poldrack et al 2011).

This lack of consensus reflects more than disagreements over the correct theory of cognition. It can be traced to the inconsistent use of conceptual terminology across tasks (Figdor 2011), and to variations in the experimental protocols used to realize phenomena in experiments (Sullivan 2009 discusses analogous challenges in neurobiology; also, see Sullivan 2016). Additionally, there is a growing consensus that, even when the meaning of the terms is agreed upon, the cognitive taxonomy is not ideal for capturing the mapping between brain processes and cognition (see Bunzl, Hanson and Poldrack 2010). The rationale for this is partly empirical, as terms and concepts used to theorize about cognitive processes stand in a many to many relationship with regional brain activity (Price and Friston 2005), and network activity (Pessoa 2014). A fact revealed by comparing findings from disparate subfields (as in Price and Friston 2005), which has since been reinforced by evidence provided by meta-analyses (Lenartowicz et al 2010; Yarkoni et al 2011). This is taken by some within the community as an indication of the failure of theories of cognition to reflect 'the brain's native ontology', and is a central motivation for efforts to revise and redefine the cognitive ontology (Anderson 2015). Data-driven revision projects, as it happens, are one of the research programs that have been made possible by the availability of large data sets and meta-analysis techniques (of which Leonartowicz et al 2010 is an example; and Klein 2012 and Anderson 2015 discuss the strategy in more detail).[20]

---

[20] By data-driven approaches, I refer to those revision projects that begin from a large data set and apply machine learning analysis tools to identify the categories and concepts that 'best' capture the patterns in the collected brain activity data. These approaches are data-driven as they aim to be agnostic about which cognitive theories are best, and instead let the available data 'decide'. Anderson's functional fingerprinting approach is one example of a data-driven approach to ontology revision. Roughly, the process involves using an algorithm that identifies the minimum set of variables that maximally

While tools like NeuroSynth are developed by members of the scientific community, and collect together neuroimaging data produced by cognitive neuroscientists for a diverse range of theoretical purposes, they are not (yet) a hub around which the community is organizing itself. This is noteworthy because, in other domains of the life sciences where data repositories have had a positive impact on knowledge production, community uptake and engagement with the repository development was an important factor in the success of these initiatives (e.g., Ankeny and Leonelli 2012; O'Malley and Soyer 2012; Leonelli 2012; Leonlli 2015). Community coordination is necessary for achieving the development of a shared taxonomy or ontology that could alleviate the barriers to data integration noted above. That is, to eliminate the terminological ambiguity and inconsistencies in conceptual and methodological practices across research programs that make data integration, and the progress it promises, difficult to achieve (see Sullivan 2017 for a detailed discussion). It is not enough to improve community coordination, and alter research practices so that they support the downstream integration of experimentally produced data into repositories. The research practices themselves, and in particular approaches to the analysis and interpretation of data, must also be adapted to the use and analysis of synthesized data sets. The methods, techniques and reasoning procedures that are applied to the analysis and interpretation of neuroimaging data are not appropriate for the interpretation of large bodies of synthesized neuroimaging data.

---

captures variation in brain activity patterns defining the 'functional fingerprint' of each region. The variables, then, are examined to determine the cognitive capacities they might refer to (see 2014, chapter 4). An alternative is the machine learning approach used by Lenartowicz and colleagues (2010). This method uses a machine learning classifier to determine which cognitive term can be discriminated between on the basis of brain activity patterns and which cannot. Those that cannot be discriminated become candidates for being removal from the ontology.

Looking to work on the epistemic character of integrative practices in other life sciences, collaboration is often emphasized as an important aspect of successful data-intensive research practices (Leonelli 2013; O'Malley and Soyer 2012; Sullivan 2017). Collaboration is important because integrating data involves "… abstracting data from their original sources…" to produce a new body of information (O'Malley and Soyer 2012, p. 61), and effectively interpreting a diverse body of data requires familiarity with the material objects that the data are about, as well as the methods and tools used to produce it (Leonelli 2013). Sabina Leonelli argues that manipulation of data such that it can be integrated with other data sets broadens its evidential scope by making it relevant for evaluating hypotheses and theories it may not have been produced to investigate (2009). However, data manipulations also restrict the scope of a data set as they inevitably suppress patterns and information in the data (Good 1983; Wright forthcoming). The processes of data manipulation used to abstract and integrate data is itself a tool that changes the evidential value of data by both expanding it and restricting it, and so familiarity with these process is also important for the effective use of synthesized data.

Data manipulations are used to eliminate noise, by suppressing the influence of detectable causal factors unrelated to the phenomena of interest — such as head motion, or a measurable drift in the magnetic field strength of the scanner. They are also used to emphasize patterns in the data, allowing human investigators to make judgements about the evidential significance of complex and multifaceted data sets. This is one of the primary functions of data analysis and manipulation in neuroimaging research: to identify patterns in a complex data set that are relevant for evaluating the hypotheses under consideration (see Wright forthcoming). Consider the use of subtraction analysis, which involves subtracting BOLD signal measurements between two task conditions, to identify parts of the brain that 'preferentially activate' for one task over another. The method itself has often been the target of criticism (Uttal 2001; Hardcastle and Stewart 2002; Klein 2010; Aktunç 2014), and has declined in popularity with the development of machine learning methods of data analysis (Haxby 2010; Kriegeskorte and Kievit 2010). However, it remains useful for addressing basic questions about regional involvement in cognitive processing. Even in the wake of powerful techniques like the machine learning

methods alluded to above, subtraction analysis results are recognized as important for clarifying the implications of more sophisticated techniques (Coutanche 2013; Davis and Poldrack 2013).

Whether to remove the influence of noise, or to emphasize aspects of the data set investigators regard as informative about a hypothesis under consideration, data manipulations support the interpretation of data by suppressing patterns and information. Specific methods of analysis and interpretation, such as subtraction or machine learning methods of analysis, are regarded as useful for answering specific questions about the data and underlying phenomena, in part because they are sensitive to different variations in the data. For instance, machine learning methods have been shown to be sensitive to variations between BOLD signal values at a voxel-by-voxel level, while univariate methods like subtraction are sensitive to variation in activity level between subjects that machine learning methods are unlikely to pick up on (Davis et al 2014). Which is to say, the implications of a data analysis result are contingent on facts about the data's production, but also on the operations that make up the analysis technique itself.

The results of a single analysis technique are rarely able to definitively discriminate between hypotheses and theories neuroimaging experiments are designed to test. Multiple techniques are used to isolate a variety of patterns, which together are used to assess the significance of the data set with respect to the hypotheses and theories under consideration (Wright forthcoming). The details of the tasks, behavioural performance, analysis techniques used to interpret the data, and their resulting data patterns, are all important in the context of the original research for drawing conclusions. These details are often suppressed in the process of data integration because they are not shared by all data sets — this is one way that data integration involves abstraction. Integrating data can restrict their evidential scope as data patterns relevant for making inferences about the phenomenon it was produced to study may be removed to facilitate the smooth integration with other data sets. To get a sense of how synthesized data sets may have different evidential value from the data sets drawn on to construct them, consider the following example.

If you search the term ‘working memory’ in NeuroSynth there is a prominent area of activation roughly centred on the coordinates X: -44, Y: 0, Z: 36. You can use NeuroSynth to find studies that report activation within a specified radius of these coordinates — these are the results that positively contribute to the forward and reverse inference maps relating these coordinates and the term ‘working memory’. One such study is the work of Todd and colleagues to identify the region of the brain that encodes visual working memory (2011). Limitations of fMRI temporal resolution make it difficult to distinguish between visual working memory encoding, perceptual and maintenance-related activity (p. 1528). To overcome this limitation Todd and colleagues had subjects perform two working memory tasks. In one task the participants had to distinguish between two faces and in another they had to distinguish between two colours. Other research has shown that encoding two faces takes almost ten times as long as encoding two colours into working memory. This allowed the researchers to distinguish visual working memory encoding from perceptual and maintenance-related processing because, “... brain regions involved in [working memory] encoding should show differential durations of activity depending on the time it takes to encode objects of different complexity” (p. 1528). The results identify the region roughly centred on the coordinates noted above as the only area of the brain where the time course of the measured signals matches this prediction.

Todd and colleagues’ conclusion is supported by a pattern in the data which is not included in the NeuroSynth database. The details of the task design and the time course of the signal that justify the inferred relation between regional brain activity and working memory are removed by NeuroSynth's automated curation procedure. The data reflected in NeuroSynth are, at best, a rough proxy for the data that supports the claim that a particular brain region is involved in encoding visual stimuli during a working memory task.[21] The details of the data set necessary for isolating patterns in the data that support

[21] This is not to say NeuroSynth has limited value. NeuroSynth is useful for identifying connections between areas of research that might not be apparent when data is considered

the claims inferred from it are potentially absent from the synthesized data set. The methods of analysis and interpretation adequate for evaluating the evidential significance of a single experimentally produced neuroimaging data set are not guaranteed to be successful when applied to the evaluation of a meta-analysis result. The synthesized data set, by virtue of the processes of manipulation required to integrate disparate data, is different in kind from the experimentally produced data sets that it is made from. This presents a risk for the promise of meta-analyses: it may not be sufficient to coordinate community practices to facilitate the integration of data. The methods of analysis and interpretation need to also be adapted to suit the use of meta-analysis tools and interpretation of synthesized data sets.

In the remainder of this paper I argue that inferences made on the basis of a meta-analysis informed by research methods honed for the interpretation of locally produced neuroimaging data are likely to be error-prone, and yet will appear to be justified from the investigator's perspective. The appearance of justification arises from the recognized soundness of the approach to data analysis and interpretation in the context of a

---

in isolation. For example, while 'working memory' is strongly correlated with those particular coordinates, the terms 'phonological', 'frontal eye' and 'saccade' are also strongly correlated with those coordinates. This suggests that there may be data from research on the phonological loop (which is a component of one psychological model of working memory, for example see Baddeley and Wilson 2002), as well as from research on the control of the visual system, relevant to the study of visual working memory. This second connection may be of particular value to Todd and colleagues as one important difference between their stimuli (which may contribute to the increased encoding time) is a difference in the way the stimuli are scanned, as defined by saccade sequences. While the stimuli and task design prevent differences in sustained focal attention from driving differences in neural activity, there was no control for differences in saccade sequences (and thus the active direction of attention) from driving the differences in activity.

neuroimaging experiment, and the errors arise due to the differences between neuroimaging data sets and the synthesized data sets available from data repositories. To make this argument, I examine a recent dispute over the evidential value of NeuroSynth data between researchers who argue that the data are evidence for the claim that a discrete region of the brain is selective for pain processing and the developer of the database who claims otherwise. I argue that the dispute, and its persistence, is due to the very problem gestured towards above: methods of data interpretation adequate for determining the value of neuroimaging data are applied to NeuroSynth data, resulting in apparently-justified, but likely mistaken, inferences.

4.4 NeuroSynth Data and Pain Selectivity

Neuroimaging data sets are, at best, able to support claims about the relationship between observed brain activity and the specific cognitive processes brought about by the cognitive tasks participants perform. This makes it difficult to use neuroimaging data to support general claims about the contribution specific regions of the brain make to the realization of cognitive phenomena. Meta-analyses are expected to overcome this limitation through the aggregation of a diverse body of data that represents a broader collection of cognitive phenomena than can possibly be brought about in a neuroimaging experiment (Kober and Wager 2010). This is what motivated Matthew Lieberman and Naomi Eisenberger (L&E hereafter) to use NeuroSynth, a repository of published neuroimaging results, to examine the relationship between activity in the dorsal anterior cingulate cortex (dACC), and the numerous cognitive processes correlated with it.

L&E's collaborative work includes the use of neuroimaging technology to investigate the neural correlates of ‘social pain’, such as rejection or exclusion from activities (Eisenberger, Lieberman, and Williams 2003; Lieberman and Eisenberger 2004). In this work they find that social rejection is correlated with activity in regions of the brain separately associated with pain (the dACC included). This has led to more recent work promoting and defending their view that social and physical pain, while phenomenologically distinct, rely on shared neural systems (Eisenberger and Lieberman 2005; Eisenberger et al 2006; Eisenberger 2015). This view is not without its challenges. One alternative is the salience account, which explains the activation of dACC in both

social and physical pain conditions by a sensitivity to highly salient stimuli, such as feelings of pain and rejection (see Eisenberger 2015). While not direct alternatives, dACC activity has also been shown to activate during memory tasks (Wager and Smith 2003), emotion (Etkin, Egner and Kalisch 2011), and a number of other cognitive processes and states that are not easily classified as 'pain' (Lieberman and Eisenberger 2015, p. 15250). Since L&E's argument that physical and social pain share a neural substrate depends on neuroimaging evidence showing that dACC is active under both conditions, these results are confounding for their view. The big problem with this domain, as with many in cognitive neuroscience, is that neuroimaging experiments cannot provide evidence for reverse inference claims. Furthermore, evidence for reverse inferences is just what is needed to discriminate between competing accounts of dACC function (Eisenberger 2015, p. 619; Berkman, Cunningham and Lieberman 2014, p. 144). This also happens to be the kind of evidence NeuroSynth has been promoted as able to provide (Yarkoni et al 2011).

In an effort to determine which of the candidate accounts of dACC functions is supported by reverse inference evidence, L&E use NeuroSynth to "... explore the best general psychological account of [dorsal anterior cingulate cortex] dACC function" (2015, p. 15250). Reporting on comparisons of reverse inference maps from NeuroSynth for terms associated with the competing accounts of dACC function, they conclude that 'pain' provides the best account of dACC function. Shortly after the paper was published, the lead developer of Neurosynth, Tal Yarkoni, thoroughly criticized the paper in a pair of blog posts (2015a; 2105b). Yarkoni's position is that "... Neurosynth data does not support any of the main claims ..." of the paper, arguing that L&E's choice of analysis methods, and interpretive decisions, amount to a misuse of NeuroSynth data (2015b). This dispute played out mostly through blog posts, with Lieberman replying to Yarkoni's first post on a separate blog (Lieberman 2015), which Yarkoni engages in his second post with a point-by-point critique. Later, a letter to the editor (Wager et al 2016), and an official reply by the authors (Lieberman et al 2016) were published, continuing the same lines of argument articulated in the blog posts. In the end, L&E were unconvinced by Yarkoni and colleagues' arguments that their analysis and interpretation of NeuroSynth data is mistaken.

This dispute is informative because it involves the use of a meta-analysis tool to conduct the very kind of research these tools are purported to enable, and yet the results instigated a heated debate between the database developer and well-intentioned users. Why did this dispute occur, what factors contributed to it, and why might L&E's unwavering position be justified? The answer, as suggested at the end of the previous section, is that L&E's approach to the analysis and interpretation of the data is not appropriately sensitive to the inferential limitations of NeuroSynth data, and yet it is also adequate for the interpretation of neuroimaging data. To argue for this position, I first review L&E's findings, and the rationale offered for their interpretation, and then Yarkoni's criticism.

4.4.1 The dACC is Selective for Pain

Lieberman and Eisenberger's most general conclusion is that "... the clearest account of dACC function is that it is selectively involved in pain-related processes" (2015, p. 15255). A region is selective for a process when there is evidence for the generalizability of both forward and reverse inferences between observations of activity in that region and instantiations of that cognitive process. As noted above, evidence that can support reverse inferences is not available from most neuroimaging experiments. To this end, Lieberman and Eisenberger justify their use of NeuroSynth by noting that it offers "… the opportunity to perform comprehensive reverse inference analyses that include virtually every psychological process that has been attributed to dACC" (p.15250).

To determine the significance of the data with respect to the claim that dACC is pain selective, they examine the forward and reverse inference maps for a collection of terms associated with four candidate psychological categories. Each of the categories considered — pain, executive control, conflict processing and salience — is matched with one to six terms in NeuroSynth (for instance, "pain", "painful" and "noxious" are the terms they used to capture the category of 'pain'). They justify the selection of four categories by identifying them as the best candidates for dACC function currently available in the relevant literature. They also provide forward inference maps from NeuroSynth as confirmatory evidence for this claim, and further justification for this decision (p. 15251).

The reported reverse inference maps show ‘pain’ terms as having a greater overall density of dACC activity than the terms associated with the other three categories. On this basis, they conclude that “… the only psychological phenomenon that can be reliably inferred given the presence of dACC activity is pain” (p. 15251). The analysis results L&E take to be the most important for their conclusion are the probability of the term occurring given activation in dACC, and the associated z-scores, which are a measure of statistical significance. L&E compare the posterior probabilities in each of eight evenly distributed coordinates from the dACC, and find that, at seven of the eight points, only the posterior probability estimates for “pain” are statistically significant. They interpret this as “... strong evidence that dACC activity in seven out of eight foci … could be attributed to pain by quantitative reverse inference” (2015, p. 15252). It is noteworthy that it is not the value of the posterior probability estimates themselves, but the statistical significance of those values, that L&E regard as the relevant criterion for assessing the evidential value of the data.

In a similar vein, L&E also compare the z-scores for pain reverse inference maps in dACC with the z-scores for reverse inference maps for all other terms in the database. They find that pain z-scores in the dACC are either the largest of all terms or, in the case of one coordinate, second only to the term ‘clinical’. They interpret this result as ruling out the possibility that terms not considered in the above analyses are better candidates for reverse inferences from dACC activity (p. 15253). It is on the basis of these two comparisons of statistical significance, that L&E come to regard the NeuroSynth data set as strong evidence for the claim that ‘pain’ is the best general psychological account of dACC processing. Yarkoni's position, on the other hand, is that NeuroSynth data provides evidence for the contrary conclusion: dACC activity and pain are not bound by a relation of selectivity. Yarkoni offers a number of arguments for this position, two of which are aimed at disputing L&E’s primary claim and are relevant to my aims here.

4.4.2 The dACC is not Selective for Pain

Two of the arguments Yarkoni offers are aimed at establishing that (1) NeuroSynth data cannot provide strong evidence for reverse inference claims, and (2) L&E's method of analysis and interpretation is not able to support their interpretive aims. Yarkoni argues

that while NeuroSynth data could provide some evidence for pain selectivity, further evidence showing that "... no other process activates dACC in a meaningful way independently of its association with pain" (Yarkoni 2015a) is required. He contends that NeuroSynth data are only able to provide weak evidence in support of reverse inferences. Part of the reason for this is that NeuroSynth data are the product of an automated curation process, which, as with any curatorial process, involves abstracting away from the details of the experiment the data are drawn from in order to integrate them seamlessly into the database. While the automated process NeuroSynth uses will retain the fact that activity in a region is correlated with an ascribed label, it does not necessarily include the aspects of the data and experimental conditions which are needed to justify the claim that the label and regional activity are correlated. The letter to the editor reinforces this point, concluding that "… Neurosynth is useful for exploring structure-to-function mappings … but it cannot provide definitive inferences about specific brain regions" (Wager et al 2016).

There is a deeper problem that Yarkoni raises for L&E's analysis: a comparison of reverse and forward inference maps over four categories does not represent a sufficiently diverse collection of candidate cognitive processes to warrant a reverse inference claim. The problem with L&E's approach is that restricting comparisons to only four categories recreates the epistemic limitations of neuroimaging experiments that NeuroSynth aims to alleviate. L&E's decision to do so is justified by reference to the extant literature on dACC function, which is the product of a focussed effort to use neuroimaging experiments to identify candidate processes for dACC function. These experiments, however, lack generalizability— they cannot be used to establish a general claim about dACC function, only a claim about dACC function relative to those cognitive capacities explicitly targeted by the tasks. This is one reason for L&E's use of the NeuroSynth database, to diversify the cognitive states represented by the data set. Unfortunately, their methodology, informed as it is by the standards of research guiding the conduct of neuroimaging experiments, and the interpretation of the resulting data, leads them to overlook patterns in the NeuroSynth data that contradict their conclusion. This problem is exacerbated by the fact that they compare the z-scores of the reverse inference maps, and not posterior probabilities. Where z-scores provide a measure of statistical significance,

posterior probabilities provide an estimate of the accuracy of a reverse inference. The latter is, contrary to L&E's view, the data pattern most relevant for evaluating the selectivity of dACC for pain.

Considering a more diverse selection of categories and comparing posterior probability estimates provides evidence that dACC is not pain selective. Wager and colleagues letter to the editor notes that "… using the same database, we estimate the probability of a study including physical pain given activity in pain-selective dACC at ~12%, on par with language, emotion, attention, and memory" (2016, p. E2474). Yarkoni, additionally, provides a comparison of posterior probability estimates for a selection of terms — motor, fear, reward, working memory — showing that, while the z-score for 'pain' is greater than that for 'motor', the probability that 'motor' is a term labelling the data given that any activity is reported in dACC is around 18%, while pain is around 8%. These results show that "… it's probably a bad idea to infer any particular process on the basis of observed activity, given how low the posterior probability estimates for most terms are going to be" (Yarkoni 2015b).

In the next section I argue that L&E's judgement of the significance of NeuroSynth data is directed and informed by a research strategy that is appropriate in the context of a neuroimaging experiment, where efforts to maximize the reliability of the data limit the generalizability of the supported claims and hypotheses. Applied to a data set arrived at by integrating a diverse body of data, their approach leads them to make a number of inferential errors — as identified by Yarkoni and colleagues. The dispute is, from this vantage point, over the appropriateness of methods for analyzing and interpreting synthesized data sets available from NeuroSynth.

4.5 The Analysis and Interpretation of Synthesized Data

There are two factors that contribute to the dispute between Yarkoni, the developer of the database, and Lieberman and Eisenberger, database users. They are (1) the application of analysis strategies appropriate in the context of neuroimaging research to the interpretation of NeuroSynth data, and (2) a misunderstanding of particular data

manipulations and the meaning of the patterns they produce. L&E's interpretation of the data is mistaken, in part, because decisions made in their analysis and interpretation of NeuroSynth data are guided by an interpretive strategy that is tuned for the interpretation of data produced in a neuroimaging experiment, and not a data set made from the integration of a diverse collection of neuroimaging data. The misunderstanding of z-scores further suggests that the understanding investigators have of the operation and meaning of data analysis techniques plays a signifiant role in directing judgements of the evidential significance of the data those techniques are applied to.

L&E focus on four specific cognitive processes because the categories of pain, executive control, conflict processing and salience are the four accounts of dACC function that are currently best supported in the literature they are contributing to (Lieberman 2015). In a review of research on social and physical pain (which Lieberman refers to in his blog post), Eisenberger argues for the view that social and physical pain share an underlying cognitive state — the "… feeling of distress or suffering and the motivation to put this experience to an end" (p. 607).[22] The alternative accounts of dACC function Eisenberger presents in that article match with the categories, and terms, L&E considered in their NeuroSynth analysis. The empirical question that this leads to is the one their NeuroSynth analysis addresses: which of pain, salience, or a number of cognitive accounts best explains dACC activity? This is why their NeuroSynth analysis takes into consideration only the terms as associated with these four cognitive processes. While it is misleading to take this approach in evaluating the evidential significance of NeuroSynth data for a selectivity claim, this approach is methodologically sound when considered in

---

[22] Eisenberger doesn't just rely on neuroimaging evidence to argue for this point. The evidence presented in support of this view largely consists in research showing that tasks inducing physical and social pain activate the dACC (e.g., Eisenberger 2003; Wager et al 2009; Kross et al 2011) as well as evidence from other areas of neuroscience, such as lesioning research correlating dulled pain experiences with damage to dACC (Eisenberger 2015, p. 604-5).

the context of a neuroimaging experiment and the more restricted inferences those experiments can be used to make.

In neuroimaging experiments the number of distinct cognitive processes comparable when analyzing and interpreting the data are limited by the details of the tasks participants performed and how they performed them. This is necessary because, without these controls, it would be (more) difficult to discern the evidential significance of neuroimaging data, given that it is influenced by causal factors other than those that are associated with the phenomena it is collected or produced to learn about. Many of these factors are idiosyncratic to the instances of measurement or data collection, and so are difficult to identify. Recognizing that data are noisy in this way entails an epistemic challenge that all scientists must overcome. That is, researchers somehow use data influenced by both the phenomena of interest and innumerable causal factors idiosyncratic to the circumstances of data production as evidence for claims about phenomena that occur in a variety of settings and circumstances (see Bogen and Woodward 1988; McAllister 1997; Woodward 2000; Harris 2003; Schindler 2011; Massimi 2011; Apel 2011 for a variety of accounts of these inferences, and arguments identifying various factors that influence them).

Experiments can be divided into two broad steps: data production and data interpretation. Following Jim Woodward's classification, data production "… has to do with the causal processes that lead from the phenomena of interest to the data", while data interpretation "… involves the use of arguments, analytic techniques, and patterns of reasoning which operate on the data so produced to reach conclusions about phenomena" (2000, p. S165). Data production has to do with what is done to create a data set, while data interpretation captures those steps involved in assessing data's evidential significance. There are two broad strategies for strengthening inferences in experimental science: improve the process of data production (i.e., 'build a better telescope', or design a better experiment), or improve the process of data interpretation (i.e., use better statistics, a new theory, or different reasoning strategies). Improving upon the process of data production results in a data set that is less noisy, and so more reliable as evidence (Woodward 2000; Sullivan 2009). The problem with L&E's inference is that, while it is based on 'better data' than a

neuroimaging experiment could provide for investigating the selectivity of a region, the data interpretation process that they use is not adequate to the task.

In designing an experiment to address an empirical question, such as the one L&E use NeuroSynth data to pursue, researchers will attempt to constrain the experimental conditions to ensure that the resulting data are able to discriminate between the specific hypotheses under evaluation. This effort aims toward the ideal data production process that Woodward describes, which is one that produces "… different sorts of data … in such a way that investigators can use such data to reliably track exactly which of the competing claims … is true" (2000, p. S166). Decisions about which processes to focus on are guided by the current literature, which will include proposals and evidence for a number of competing hypotheses that experiments can be designed to discriminate between (see also Sullivan 2009; 2016). There is a tradeoff in using experimental controls to enhance the capacity of data to discriminate between a specific set of hypotheses (Sullivan 2009; 2015).

Jacqueline Sullivan argues that the reliability of data and their capacity to warrant inferences that are true of phenomena produced outside of laboratory settings are in tension because the controls necessary to improve the reliability of data often involve creating circumstances far removed from those that occur outside of laboratory settings. Research in neurobiology emphasizes reliability, and as a result "… inevitably restricts the extension of interpretive claims to the laboratory" (Sullivan 2009, p. 535). The same could be said of neuroimaging experiments (see Sullivan 2015). Indeed, this tension captures part of the problem with using neuroimaging data to conduct reverse inferences.

Reverse inference claims, such as 'dACC activity indicates pain-related processing' are about the cognitive capacity, state, or process indicated by the engagement of a region of the brain. A data set that could provide this evidence, at minimum, must represent as many cognitive states and processes as dACC could conceivably be involved in bringing about. Since many correlations between regional activity and cognitive processes go unreported, these processes must include, but not be limited to, those that dACC has specifically been implicated in. Neuroimaging experiments, however, prioritize reliability

with respect to discriminating between a specific set of hypotheses, and so cannot provide the requisite evidence to support a generalizable reverse inference (see Poldrack 2006; Klein 2012; Machery 2014).

L&E regard NeuroSynth data as valuable for overcoming the limitations of neuroimaging experiments because the analyses it provides include "… virtually every psychological process that has been attributed to dACC" (2015, p. 15250). This, however, only goes part of the way to overcoming the problems with reverse inferences. While a data representing a broad selection of cognitive processes are necessary to support a reverse inference claim, simply possessing data that could be used as evidence for a reverse inference claim is not sufficient to warrant such a claim. One one hand, the tradeoff between external validity and reliability cuts both ways. Synthesizing data sets to secure greater external validity for reverse inference claims means reducing the reliability of the data with respect to those very claims. The process of data synthesis also eliminates patterns in the data that are important for determining what each data set was originally about (see the example at the end of section 3, and also Sullivan 2016). This is particularly true for NeuroSynth, as its curatorial process is implemented by an automated computational procedure. This is why Yarkoni insists that NeuroSynth data, at best, can support weak claims. NeuroSynth data cannot stand on their own as evidence warranting a reverse inference, let alone a claim about the selectivity of a region for a cognitive process. The other reason data aggregation is not sufficient as a solution for the problems of reverse inference is that the problems go beyond problems with data production processes. Many methods of data interpretation and analysis popular in neuroimaging research are unable to provide clear indications of whether or not a reverse inference claim is supported by the data (Poldrack 2010, p. 755-6).

Proposals for improving the quality of reverse inferences typically recommend improvements to both data production and data interpretation procedures. Colin Klein, in addition to recommending a focus on networks over regions, gestures towards meta-analyses as the way forward (2012); Russell Poldrack promotes machine learning methods of data-analysis as a tool for conducting investigations of reverse inference claims using large bodies of synthesized data (2012); and Clark Glymour and Catherine

Hanson propose a different analysis strategy from Poldrack, which also depends on the availability of a large and diverse body of neuroimaging data (2016). Each of the proposals recommends that both better data and a new method of analysis are needed to assess the data's evidential significance with respect to reverse inference claims. L&E also use a new method of analysis to evaluate the data, relying on the reverse inference maps which are produced by a machine learning method similar to the one Poldrack argues formally implements reverse inferences (2011; see also Yarkoni et al 2011). The problem with L&E's approach is that they've only implemented a change in the data and analysis techniques, and not a corresponding change in their approach to the data's interpretation.

By restricting their focus to the four categories identified within the specific literature they are contributing to, L&E are assuming that the diversity in the cognitive states reflected in the data set is a sufficient change to their research practices for evaluating a selectivity claim. This would be fine if L&E were analyzing data produced in an adequately controlled experiment, and if they were content to restrict the inferred claims accordingly. As presented, however, they draw the stronger conclusions that dACC is pain selective. Their approach to evaluating the evidential significance of the data with respect to this conclusion, that is the process of data analysis and interpretation they use, is not up to this task.

This dispute reflects a conflict between approaches to the analysis and interpretation of data appropriate in the context of a neuroimaging experiment and approaches appropriate for evaluating the significance of a synthesized data set. The analysis strategy L&E apply, while it may be misleading when used to evaluate NeuroSynth data, is justifiable in the context of a neuroimaging experiment. In this way, L&E's interpretation can be regarded as justified, if misleading. This presents a problem for the prospects of meta-analyses, above and beyond challenges to data integration that follow from the lack of coordination in methodological and conceptual practices emphasized by other commenters on the prospects of meta-analyses (Yarkoni et al 2011; Poldrack et al 2017; Sullivan 2016; 2017). The problem, as it relates to the misuse of powerful new technologies for data analysis and interpretation, is not unique to meta-analyses, but is a

problem for cognitive neuroscience more generally. The general problem is related to the second problem with L&E's interpretation: it is based on a misunderstanding of data analysis techniques and the meaning of the data patterns they isolate.

The perception of tools like NeuroSynth as 'solving' the problems with neuroimaging experiments with respect to reverse inferences is part of what builds L&E's confidence in the soundness of their approach — they say as much in the introduction of their paper. This is not to say L&E simply trust that NeuroSynth data are adequate to this task. They consider some of the facts of NeuroSynth's curatorial process, as they note that "... the reverse inference is linguistic, focused on the terms used across articles rather than on task trial types of specific psychological states" (2015, p. 15254), and in their official reply to Wager and colleagues, they report that they manually verified the ascription of a random selection of labels to fifty papers for each category (2016). Even so, L&E still mistake it as providing strong evidence for the claim that dACC is pain selective in part because they understand NeuroSynth as a tool that can provide evidence sufficient for definitively evaluating a reverse inference claim. A similar misunderstanding affects their assessment of NeuroSynth's evidential significance and guides the decision to focus on z-scores associated with the considered reverse inference maps.

In justifying the focus on z-scores over posterior probabilities, Lieberman explains that they were "not interested in effect sizes" (which is what posterior probabilities are), but wanted to evaluate the "accumulated evidence for the reliability of reverse inferences" and for this reason focussed on z-scores (2015). If z-scores reflect the accumulated evidence for the reliability of an inference, then a higher z-score indicates a more likely reverse inference target. Whether or not L&E's interpretation of z-scores is accurate, that they understand them as they do explains why they judge the data to be positive evidence for the pain-selectivity of dACC.[23] Yarkoni, on the other hand, is familiar with the details

---

[23] Comparing the z-scores for reverse inference maps associated with different terms assumes that differences between z-scores are statistically significant. This is an assumption that has been criticized by statisticians on the grounds that differences in

of the NeuroSynth algorithm and meta-analysis tools, having developed them. Yarkoni rightly regards z-scores as estimates of statistical significance, since they are computed by transforming p-values, and views effect sizes as the relevant statistical output to compare when evaluating selectivity claims. Posterior probabilities, and not z-scores, are the relevant data pattern for judging the selectivity of a pattern of brain activity for a cognitive process or state.

Tools for the efficient conduct of large-scale meta-analyses, like NeuroSynth, are only the latest example of novel techniques for manipulating and analyzing data driving progress in cognitive neuroscience. The last two decades of neuroimaging research has seen the development and uptake of a diverse collection of distinct tools and technologies for manipulating and analyzing data. This has been viewed as both valued and dangerous - with neuroscientists cautioning that with the increasing number of methods for analyzing data comes an increased rate of false positives (Carp 2012a; 2012b), an increased potential for misinterpretation of results, more opportunities for questionable research practices, and the potential of 'getting lost in data' (Poldrack et al 2017; Munafò et al 2017). In response to some of these worries, Eve Marder has argued that, as data analysis processes become more complex and more diverse, intuitions about how to

---

statistical significance are often not themselves significant (Gelman and Stern 2006), and has been identified as an ubiquitous and problematic practice in the neuroimaging research (Nieuwenhuis et al 2011). Lieberman's response shows that they are unaware of this problem, as they interpret a difference in z-scores as showing that "… we can be more confident that there is some real association between pain and dACC than between the other three terms and dACC" (2016). L&E's understanding of z-scores leads to inferring a descriptive claim about the data that is incorrect: that the evidence in support of a reverse inference from dACC activity to pain is stronger than for other terms. This in turn leads to explaining the z-score comparison by appeal to the selectivity of dACC for pain.

engage with data will become more important for driving progress in the neurosciences in a positive direction (Marder 2015). Intuitions about how to engage with, analyze and interpret data are also important because, as the dispute and discussion above shows, they can promote inferential errors and lead to faulty judgements of data's significance on the basis of interpretive strategies that are sound in other contexts. Progress requires more than just making tools, making them available, and community coordination. Open discussions about the effective use of these tools, and changes in the standards of evidence guiding the conduct and evaluation of approaches to the analysis and interpretation of data must accompany these technologies. The debate between Yarokni and Lieberman that unfolded in the blogosphere is an example of an open discussion, but such discussions also need to be channeled into changes in research practices if they are to be productive.

Disputes between stake holders in the scientific community are all but guaranteed in the context of data sharing and meta-analysis, in part because individuals, labs, and research groups have different goals, theoretical backgrounds, technical expertise and research priorities. One striking difference between the data sharing efforts in cognitive neuroscience and those operative in other biological sciences is that cognitive neuroscientists have not instituted a mechanism of 'governance'. In the case of bio-ontologies — which are the taxonomic structures used to label, categorize and organize data within a database — consortia were formed that effectively forced curators, regulators and users of databases to coordinate and interact by providing a platform for stakeholders to engage in discussion, and mechanisms for dissent to be transformed into action (see Leonelli 2016, chapter 2). Such a mechanism allows disputes to be productively funneled into changes in research practices, repository structure and curatorial procedures.

In cognitive neuroscience, there are informally organized communities of like-minded researchers, database developers and curators who actively interact with users (Yarkoni, for instance, is active on a google group for NeuroSynth), but there is not yet an established regulatory body or disciplinary framework for assessing, guiding and formalizing research practices as the discipline beings to engage with and use large data

repositories. The situation in neuroimaging research may just be a temporary symptom of the growing pains associated with efforts to grow and enrich a research community through the integration of data. Cognitive neuroscience, as is to be expected, faces its own unique set of challenges including ongoing efforts to establish standardized data formats and incentivize their uptake by the wider community (e.g., Gorgolewski et al 2015), research and funding incentives that are not aligned with data sharing, and a limited allocation of resources to the sustainment of data repositories (Poldrack and Gorgolewski 2015). Even so, the argument above demonstrates that, in addition to community-wide collaboration and coordination as others have argued is necessary (Sullivan 2017), data sharing practices also need to be accompanied by changes in the standards appealed to in the assessment of the adequacy of analysis methods and interpretive strategies if they are to promote the generation of knowledge as their advocates promise.

4.6 Conclusion

The capacity to engage in large scale data integration, and conduct meta-analyses over such a database, is a relatively new possibility for cognitive neuroscientists. It remains to be seen if they will be successful at improving knowledge of human cognition in the way that database developers hope.

The integration of data into shared and accessible repositories, and the meta-analysis techniques they make possible promise to improve theories and knowledge production in cognitive neuroscience. If meta-analyses are to productively move the discipline forward, standards guiding the analysis and interpretation of data must be changed. Explicitly outlining the pitfalls and limitations of databases and meta-analysis tools is not sufficient to offset the potential application of analysis and interpretive strategies that are sound in the context of neuroimaging experiments, but insensitive to the limitations of synthesized data. Indeed, L&E refer to Yarkoni's own work outlining the limitations and effective uses of NeuroSynth as justification for their methods — which I have shown are, by Yarkoni's own lights, ill-suited to the interpretation of NeuroSynth data. They identify a paper outlining the value of NeuroSynth for reverse inference claims (Yarkoni et al 2011), as well as discussions in the google group dedicated to answering questions about

the use of NeuroSynth, as templates upon which their analysis and interpretation is based (Lieberman and Eisenberger 2015). Furthermore, rigid guidelines for data use may undermine one of the known epistemic advantages of data sharing. Sharing data promotes knowledge, in part, by creating the potential for new discoveries to be made with old data (e.g., Leonelli 2009, 2015; O'Malley and Soyer 2012). Arguments in cognitive neuroscience for sharing data often refer to this potential as justification for the allocation of resources to database development and maintenance (Van Horn and Gazzaniga 2013). It is important that investigators are given some latitude in the way they access, analyze and interpret shared bodies of data.

With respect to the broader philosophical impact of this discussion, the situation in cognitive neuroscience is relevantly different from that in other areas of the life sciences where data repositories have been successful. Neuroscientists are responding to typical challenges for open science and sharing data in novel ways. NeuroSynth is a prime example: it uses automated curation to bypass the need for the wider community to commit to data sharing in order to effectively populate a large-scale database. It is, thus, an occasion and a time in which philosophers might learn from carefully examining how efforts to integrate neuroimaging data interface with the organization of the research community. One promising approach is to compare the epistemic and social dimensions of successful data sharing practices with the current practice of cognitive neuroscience and use that as a lens through which to identify likely pitfalls and barriers to progress, as I have begun to do here. In this way, philosophical contributions may also be able to productively impact cognitive neuroscience by identifying similarities and differences between strategies proven successful in other fields and the social, empirical and practical conditions that allowed those strategies to succeed.

References

Anderson, M. L. [2015]: 'Mining the Brain for a New Taxonomy of the Mind', Philosophy Compass, 10, pp. 68-77.

Anderson, M. L. [2014]: After Phrenology: Neural Reuse and the Interactive Brain. MIT Press.

Apel, J. [2011]: ‘On the Meaning and the Epistemological Relevance of the Notion of a Scientific Phenomenon’, Synthese, 182(1), pp. 23-38.

Baddley, A., and Wilson, B. A. [2002]: ‘Prose recall and amnesia: implications for the structure of working memory’, Neurospychologia, 40(10), pp. 1737-1743.

Berkman, E. T., Cunningham, W. A., and Lieberman, M. D. [2014]: ‘Research methods in social and affective neuroscience’ in H. T. Reis & C. M. Judd (Eds.) Handbook of research methods in personality and social psychology (2nd ed), (pp. 123-158). New York: Cambridge University Press

Bogen, J. and Woodward J. [1988]: ‘Saving the Phenomena’, Philosophical Review, 97, pp. 303-52.

Bunzl, M., Hanson, S. J., and Poldrack, R. A. [2010]: ‘An exchange about localism’ in M. Bunzl and S. J. Hanson (eds), Foundational Issues in Human Brain Mapping, The MIT Press.

Carp, J. [2012a]: ‘On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments’, Frontiers in Neuroscience, 6, pp. 149.

Carp, J. [2012b]: ‘The secret lives of experiments: Methods reporting in the fMRI literature’, NeuroImage, 63(1), pp. 289-300. DOI: 10.1016/j.neuroimage.2012.07.004

Carter, C. S., Botvinick, M. M., and Cohen, J. D. [1999]: ‘The contribution of the anterior cingulate cortex to executive processing in cognition’, Rev Neuroscience, 10(1), pp. 49-57.

Costafreda, S. G. [2011]: ‘Meta-Analysis, Mega-Analysis, and Task Analysis in fMRI Research’, Philosophy, Psychiatry, and Psychology, 18(4), pp. 275-277.

Coutanche, M. N. [2013]: ‘Distinguishing multi-voxel patterns and mean activation: Why, how, and what does it tell us?’, Cognitive Affective Behavioural Neuroscience, 13, 667-673.

Davis, T. and Poldrack, R. A. [2013]: ‘Measuring neural representations with fMRI: practices and pitfalls’, Annals of the New York Academy of Sciences, 1296, pp. 108-34.

Davis, T., LaRocque, K. F., Mumford, J. A., Norman, K. A., Wagner, A. D., and Poldrack, R. A. [2014]: ‘What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis’, Neuroimage, 97, pp. 271-83.

Eisenberger, N. I. [2015]: ‘Social pain and the brain: Controversies, Questions, and Where to Go from Here’, Annual Review of Psychology, 66, pp. 601-629.

Eisenberger, N. I., & Lieberman, M. D. [2005]: ‘Broken hearts and broken bones: The neurocognitive overlap between social pain and physical pain’, in K. D. Williams, J. P. Forgas, & W. von Hippel (Eds.), The Social Outcast: Ostracism, Social Exclusion Rejection, and Bullying (pp. 109-127). New York: Cambridge University Press

Eisenberger, N. I., & Lieberman, M. D. [2004]: 'Why rejection hurts: a common neural alarm system for physical and social pain', Trends in Cognitive Sciences, 8, pp. 294-300.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. [2003]: 'Does rejection hurt? An fMRI study of social exclusion', Science, 302, pp. 290-292.

Eisenberger, N. I., Jarcho, J. M., Lieberman, M. D., & Naliboff, B. D. [2006]: 'An experimental study of shared sensitivity to physical pain and social rejection', Pain, 126, pp. 132-138.

Etkin, A., Egner, T., and Kalisch R. [2011]: 'Emotional processing in anterior cingulate and medial prefrontal cortex', Trends in Cognitive Science, 15(2), pp. 85-93.

Figdor, C. [2011] 'Semantics and Metaphysics in Informatics: Toward an Ontology of Tasks', Topics in Cognitive Science, 3(2), pp. 222-226.

Glymour, C., and Hanson, C. [2016]: 'Reverse Inference in Neuropsychology',The British Journal for the Philosophy of Science, 67(4), pp. 1139-1153.

Good, I. J. [1983]: 'The philosophy of exploratory data analysis', Philosophy of science, 50(2), pp. 283-295.

Gorgolewski, K. J., Poline, J-B., Keator, D. B., Nichols, B. N., Auer, T., Craddock, R. C., Flandin, G., Ghosh, S. S., Sochat, V. V., Rokem, A., Halchenko, Y. O., Hanke, M., Haselgrove, C., Helmer, K., Maumet, C., Nichols, T. E., Turner, J. A., Das, S., Kennedy, D. N., Poldrack, R. A. [2015]: 'Brain Imaging Data Structure - a new standard for describing and organizing human neuroimaging data', Frontiers in Neuroscience, doi: 10.3389/conf.fnins.2015.91.00056

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V.V., Nichols, T.E., Poldrack, R.A., Poline, J.B. & Yarkoni, T. [2015]: 'NeuroVault. org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain', Frontiers in neuroinformatics, 9, 8.

Hardcastle, V. G., & Stewart, C. M. [2002]: 'What do brain data really show?', Philosophy of Science, 69(S3), pp. S72-S82.

Harris, T. [2003]: 'Data models and the acquisition and manipulation of data', Philosophy of Science, 70, pp. 1508-1517.

Haxby, J. V. [2010]: 'Multivariate Pattern Analysis of fMRI data' in M. Bunzl and S. J. Hanson (eds), Foundational Issues in Human Brain Mapping, The MIT Press, pp. 55-68.

Horwitz, B. [2014]: 'The elusive concept of brain network Comment on "Understanding brain networks and brain organization" by Luiz Pessoa', Phys Life Rev, 11(3), pp. 448-451.

Kober, H., and Wager, T. D. [2010]: 'Meta-analysis of neuroimaging data', WIREs Cognitive Science, 1, pp. 293-300. doi:10.1002/wcs.41

Klein, C. [2014]: 'Review: Engineering the Next Revolution in Neuroscience by Alcino J. Silva; Anthony Landreth; John Bickle', Philosophy of Science, 81(3), pp. 486-489.

Klein, C. [2012]: 'Cognitive Ontology and Region- versus Network-Oriented Analyses', Philosophy of Science, 79(5), pp. 952-960.

Klein, C. [2010]: 'Images are not the evidence in Neuroimaging', British Journal for the Philosophy of Science, 61, pp. 265-78.

Kriegeskorte, N. and Kievit, R. A. [2013]: 'Representational geometry: integrating cognition, computation, and the brain', Trends in Cognitive Sciences, 17, pp. 401-12.

Kross, E., Berman, M. G., Mischel, W., Smith, E. E., and Wager, T. D. [2011]: 'Social rejection shares somatosensory representations with physical pain', PNAS, 108, pp. 6270-6275.

Laird, A. R., Eickhoff, S. B., Fox, P. M., Uecker, A. M., Ray, K. L., Saenz, J. J., McKay, D. R., Bzdok, D., Laird, R. W., Robinson, J. L. & Turner, J. A. [2011]: 'The BrainMap strategy for standardization, sharing, and meta-analysis of neuroimaging data', BMC research notes, 4(1), 349.

Lenartowicz, A., Kalar, D. J., Congdon, E., and Poldrack, R. A. [2010]: 'Towards an ontology of cognitive control', Topics in Cognitive Science, 2(4), pp. 678-692.

Leonelli, S. [2016]: Data-Centric Biology. University of Chicago Press.

Leonelli, S. [2013]: 'Classificatory theory in biology', Biological Theory, 7(4), pp. 338-345.

Leonelli, S. [2012]: 'Classificatory theory in data-intensive science: The case of open biomedical ontologies', International Studies in the Philosophy of Science, 26(1), pp. 47-65.

Leonelli, S. [2009]: 'On the locality of data and claims about phenomena', Philosophy of Science, 76(5), pp. 737-749.

Leonelli, S., and Ankeny, R. A. [2012]: 'Re-thinking organisms: The impact of databases on model organism biology', Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43(1), pp. 29-36.

Lieberman, M. D. [2015]: 'Comparing Pain, Cognitive, and Salience accounts of dACC'. Available at: https://www.psychologytoday.com/blog/social-brain-social-mind/201512/comparing-pain-cognitive-and-salience-accounts-dacc

Lieberman, M. D., and Cunningham, W. A. [2009]: 'Type I and Type II error concerns in fMRI research: re-balancing the scale', Social Cognitive and Affective Neuroscience, 4, pp. 423-428. doi:10.1093/scan/nsp052

Lieberman, M. D., and Eisenberger, N. I. [2015]: 'The dorsal anterior cingulate cortex is selective for pain: Results from large-scale reverse inference', PNAS, 112(49), pp. 15250-15255.

Lieberman, M. D., Burns, S. M., Torre, J. B., and Eisenberger, N. I. [2016]: 'Reply to Wager et al.: Pain and the dACC: The importance of hit rate-adjusted effects and posterior probabilities with fair priors', PNAS, 113, p. E2476-E2479.

Machery, E. [2014]: 'Significance Testing in Neuroimagery', in Kallestrup J., and Sprevak, M.(eds.), New Waves in the Philosophy of Mind, Palgrave Macmillan, pp. 262-277.

Marder, E. [2015]: 'Understanding Brains: Details, Intuitions, and Big Data', PLoS Biology, 13(5), e1002147.

Massimi, M. [2011]: 'From data to phenomena: a Kantian Stance', Synthese, 182, pp. 101-116.

McAllister, J. [1997]: 'Phenomena and patterns in data sets', Erkenntnis, 47(2), pp. 217-228.

Menon, V., and Uddin, L. Q. [2010]: 'Saliency, switching, attention and control: a network model of insula function' Brain Structure and Function, 214, pp. 655-667.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E., Ware, J. J., and Ioannidis, J. P. A. [2017]: 'A manifesto for reproducible science', Nature Human Behaviour, 1.

O'Malley, M. A., and Soyer, O. S. [2012]: 'The roles of integration in molecular systems biology', Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43(1), pp. 58-68.

Pessoa, L. [2014]: 'Understanding brain networks and brain organization', Phys Life Rev, 11(3), pp. 400-435.

Poldrack, R. A. [2011]: 'Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding', Neuron, 72(5), pp. 692-697.

Poldrack, R. A. [2010]: 'Mapping mental function to brain structure: how can cognitive neuroimaging succeed?', Perspectives on Psychological Science, 5(6), pp. 753-761.

Poldrack, R. A. [2006]: 'Can cognitive processes be inferred from neuroimaging data?', Trends in Cognitive Sciences, 10(2), pp. 59-63.

Poldrack, R. A., and Gorgolewski, K. J. [2015]: 'OpenfMRI: Open sharing of task fMRI data', NeuroImage, 144, pp. 259-261.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò M. R., Nichols, T. E., Poline, J., Vul, E., and Yarkoni, T. [2017]: 'Scanning the horizon: towards transparent and reproducible neuroimaging research', Nature Reviews Neuroscience, 18, pp. 115-126.

Poldrack, R. A., Barch, D. M., Mitchell, J. P., Wager, T. D., Wagner, A. D., Devlin, J. T., Cumba, C., Koyejo, O., and Milham, M. P. [2013]: 'Towards open sharing of task-based fMRI data: the OpenfMRI project', Frontiers in Neuroinformatics, 7(12). doi:10.3389/fninf.2013.00012

Poldrack, R.A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D. S., Sabb, F. W., and Bilder, R. M. [2011]: 'The Cognitive Atlas: Towards a Knowledge Foundation for Cognitive Neuroscience', Frontiers in Neuroinformatics, 5(17). doi:10.3389/fninf.2011.00017

Price, C. J., and Friston, K. J. [2005]: 'Functional ontologies for cognition: The systematic definition of structure and function', Cognitive Neuropsychology, 22, pp. 262-275.

Reid, A.T., Bzdok, D., Genon, S., Langner, R., Müller, V. I., Eickhoff, C. R., Hoffstaedter, F., Cieslik, E. C., Fox, P. T., Laird, A.R., Amunts, K., Caspers, S., Eickhoff, S.B. [2015]: 'ANIMA: A data-sharing initiative for neuroimaging meta-analyses', Neuroimage, doi: 10.1016/j.neuroimage.2015.07.060

Schindler, S. [2011]: 'Bogen and Woodward's Data-Phenoman Distinction, Forms of Theory-Leadenness, and the Reliability of Data', Synthese, 182, pp. 39-55.

Silva, A. J., Landreth, A., and Bickle, J. [2013]: Engineering the Next Revolution in Neuroscience. Oxford University Press.

Sullivan, J. A. [2017]: 'Coordinated Pluralism as a Means to Facilitate Integrative Taxonomies of Cognition', Philosophical Explorations, 2, pp. 129-145.

Sullivan, J. A. [2016]: 'Stabilizing Constructs through Collaboration across Different Research Fields as a Way to Foster the Integrative Approach of the Research Domain Criteria (RDoC) Project', Frontiers in Human Neuroscience, 10.

Sullivan, J. A. [2015]: 'Experimentation in Cognitive Neuroscience and Cognitive Neurobiology' In The Handbook of Neuroethics (Springer), Jens Clausen and Neil Levy (Eds)., Dordrecht: Springer, pp. 31-47.

Sullivan, J. A. [2009]: 'The multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience', Synthese, 167(3), pp. 511-539.

Todd J. J., Han S. W., Harrison S., & Marois R. [2011]: 'The neural correlates of visual working memory encoding: A time-resolved fMRI study', Neurpsychologia, 19, pp. 1527-1536. doi: 10.1016/j.neuropsychologia.2011.01.040

Turner, J. A., and Laird, A. R. [2012]: 'The Cognitive Paradigm Ontology: Design and Application', Neuroinformatics, 10(1), pp. 57-66.

Uttal, W. [2001]: The New Phrenology, The MIT Press.

Van Horn, J. D., and Gazzaniga, M. S. [2013]: 'Why share data? Lessons learned from the fMRIDC', Neuroimage, 82, pp. 677-682.

Wager, T. D., and Smith, E. E. [2003]: 'Neuroimaging studies of working memory: A meta-analysis", Cognitive Affective Behavioural Neuroscience, 3(4), pp. 255-274.

Wager, T. D., Atlas, L. Y., Botvinick, M. M., Chang, L. J., Coghill, R. C., Davis, K. D., Ianetti, G. D., Poldrack, R. A., Shackman, A. J., and Yarkoni, T. [2016]: 'Pain in the ACC?', PNAS, 113, pp. E2474-E2475.

Wright, J. W. [Forthcoming]: 'The Analysis of Data and the Evidential Scope of Neuroimaging Results', British Journal for the Philosophy of Science.

Yarkoni, T. [2015a]: 'Still not selective: comment on comment on comment on Lieberman and Eisenberger (2015)'. Available at: http://www.talyarkoni.org/blog/2015/12/14/still-not-selective-comment-on-comment-on-comment-on-lieberman-eisenberger-2015/

Yarkoni, T. [2015b]: 'No, the dorsal anterior cingulate cortex is not selective for pain: comment on Lieberman and Eisenberger (2015)'. Available at: http://www.talyarkoni.org/blog/2015/12/05/no-the-dorsal-anterior-cingulate-is-not-selective-for-pain-comment-on-lieberman-and-eisenberger-2015/

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. [2011]: 'Large-scale automated synthesis of human functional neuroimaging data', Nature Methods, 8, pp. 665-670.

Chapter 5

5 Data Analysis in Neuroimaging

5.1 Introduction

The development and use of new techniques for creating, storing, transporting, transforming, analyzing and/or organizing data has been instrumental in moving cognitive neuroscience forward. These developments have brought to light novel hypotheses and phenomena that could not be investigated before the development of new techniques for creating, handling and manipulating neuroimaging data. These innovations and the progress they bring are not without their challenges. Data analysis involves manipulating and changing data through the application of sophisticated computational processes. This creates opportunities for results to be misinterpreted, or techniques to be misapplied. The rapid rate of innovation in the methods of analysis and tools for handling and sharing data, when not accompanied by equally rapid adaption and updating of research methods and practices, also creates opportunities for well-intentioned researchers to misuse and misinterpret data's evidential significance. These are reasons for the philosophical community working on the nature of evidence in neuroscience to turn an eye towards data analysis. A first step towards this end is understanding what role data analysis techniques play in cognitive neuroscience, and how that role is played. This has been the broad aim that unites the preceding three papers.

Often, when we discuss data analysis and manipulation we think of statistical tests and correcting for noise and artifacts. While data analysis techniques and manipulations are used to quantify the evidential relation between data and a hypothesis, to eliminate detectable artifacts, and minimize random noise, these are not the only functions performed by the processes and techniques used to analyze neuroimaging data. The techniques at the heart of neuroimaging research, including MVPA techniques like pattern classification analysis, make distinct contributions to the data interpretation process. MVPA techniques have, among other things, made possible a number of novel

and promising research projects, including: interrogating the validity of concepts used to describe cognitive capacities and processes (as in Lenartowicz et al 2010; Anderson 2015), tracking changes in information as it moves through the brain (as in Kohler et al 2013; Tambini et al 2013), and decoding the contents of brain activity by predicting participant behaviour (as reviewed by Tong and Pratte 2012). In chapter three I argued that these techniques are valuable because the results of data analysis techniques, unlike the 'raw' data provided by neuroimaging experiments, can be explained in terms of claims about phenomena. Data analysis techniques are useful because they isolate patterns in the data that are explicable in terms of claims about the phenomenon neuroscientists are interested in. They are, however, themselves inferentially limited. In the second chapter, I showed that multiple analysis techniques are used together to determine the evidential significance of neuroimaging data. In the fourth chapter, I argued that research goals and background theory inform the decisions made in the application of data analysis techniques, and in turn the explanations offered for the data patterns. In this last chapter I use the arguments and cases examined in the previous papers to articulate how data analysis techniques facilitate data interpretation in terms of the interpretive and epistemic leverage they provide.

To this end, the next section clarifies the distinction between data and data patterns that has been a central part of how I think about data analysis. I draw this distinction by virtue of the advantages data patterns have in comparison to the challenges with interpreting the data they are derived from. In the third section, I examine the epistemic advantages and disadvantages of multiple analysis techniques more closely — reflecting on the examples discussed throughout the preceding papers. I argue that the evidence provided by multiple data patterns is unlikely to produce genuinely robust results, as suggested at the end of the second chapter. I argue that it is not the independence or convergence of data patterns that allows a variety of them together to enhance the value of neuroimaging data, but their distinctiveness. In the final section, I reflect on the challenges and questions raised by this project given the current trajectory of neuroimaging research.

5.2 Evidence in Neuroimaging

I have argued throughout the preceding chapters that the argumentative strategy skeptics of neuroimaging research rely on is not sensitive to a number of epistemically relevant features of research practices operative in cognitive neuroscience. In the second chapter, I contributed my own defence of neuroimaging research, arguing that the skeptical strategy mistakenly treats data analysis techniques in isolation of each other (Wright forthcoming). Adequately treating of inferences in neuroimaging research requires attending to the diversity of analysis techniques used to interpret the data. It also requires recognizing and evaluating the impact that each individual technique, and decisions about which techniques to use and how to use them has on the evidential value of neuroimaging data. In the third chapter, I noted that the challenges raised by skeptics, while based on an artificial treatment of the data interpretation process, are still legitimate challenges that inferences in neuroimaging research must overcome. That is, analysis techniques require assumptions to be made of the data that can undermine the inferences they are used to support. The indirect relationship between neuroimaging data and the phenomena it is used to study, combined with the sophistication of data analysis processes further complicates this situation by making it relatively easy to identify viable alternative hypotheses. To overcome these challenges, the interpretive process must (1) not be weakened by assumptions implicit in each of its parts, and (2) provide resistance against alternative and competing hypotheses not explicitly considered in the interpretive process. This, I argued in chapter 3, is achieved through the explanation of a variety of data analysis results.

In this section I argue that data and data patterns can be distinguished by the different evidential roles that they play. This distinction serves two purposes: (1) it makes clear why the argumentative strategy of examining a single, if salient and significant, analysis technique fails, and (2) it shows how the use of multiple analysis techniques can address the two challenges noted above. This distinction is then taken up in the next section, which re-examines the epistemic advantage provided by the use of multiple analysis techniques.

5.2.1 Data, Data Patterns, and Phenomena

Recall that, on Bogen and Woodward's view (1988), data are the product of an experiment. They characterize data as "... idiosyncratic to particular experimental contexts, and typically cannot occur outside of those contexts" (317). Phenomena, on the other hand ".... have stable, repeatable characteristics which will be detectable by means of a variety of different procedures, which may yield quite different kinds of data" (317). Bogen and Woodward further argue that "... facts about data and facts about phenomena differ in what they serve as evidence for (claims about phenomena versus general theories)" (306). Data are evidence for claims about phenomena, which in turn provide evidence for the general theories that explain them. Treating data as evidence for a claim about a regularity in the world is no small task, often requiring the careful construction and implementation of an experimental circumstance that instantiates an instance of the target phenomenon such that it is susceptible to measurement. Indeed, Woodward argues that improvements to the reliability of data-phenomena inferences often comes from improvements to the quality of and knowledge about the means of data production (2000).

Woodward contrasts improvements to the quality of data production with improvements to the reasoning processes involved in data interpretation, insisting that, instead of by changes in background theory and data interpretation, "… in real science the most effective improvements in reliability very often are achieved by altering the data production process—by building a better telescope …" (2000, p. S165). As the preceding cases and arguments show, in neuroimaging research at least, significant progress has been made not only by altering the data production process (neuroimaging technology has certainly improved over the last two decades), and providing better and more sophisticated theories, but also through the development and uptake of new tools for analyzing, organizing, accessing and manipulating data. In situating data analysis as part of the interpretation process, I have been implicitly arguing that the progress made through innovations in data analysis constitute an improvement to the process of data interpretation. Variations in BOLD signal measurements and behavioural data are causally distant from the neural activity and cognitive processes investigators use those

variations to make inferences about. This extra challenge, above and beyond those that are characteristic of all experimental science, necessitates the need to improve reasoning processes. This is necessary because neuroscientists using neuroimaging data to make inferences to claims about phenomena must not only contend with noise in the data, but also the fact that the objects of measurement are indirectly related to the causal factors that give rise to the phenomena of interest.

Estimating the portion of the BOLD signal measurement that corresponds with task related activity is an example of data manipulations used to improve the evidential significance of neuroimaging data. This typically involves picking a model for the hemodynamic response function (hrf), and, as discussed in the second chapter, using deconvolution to partition the BOLD measurements into the hrf and noise. There are a variety of approaches to estimating the hemodynamic response from the BOLD signal, each of which involves making different assumptions about the causal factors involved in its production. This is why the choice of model has a significant impact on the resulting parameter estimates (see Lindquist et al 2009).

The deconvolution process plays a different epistemic role than analysis techniques like subtraction and multivariate pattern analysis. Calculating an hrf is a pre-processing step conducted to improve the strength of evidence provided by the results of subsequent analyses. In the case considered in the second chapter, the process is used to minimize the possibility that a machine learning classifier's accuracy is due to convenient correlations between biological processes that are concurrent with, but irrelevant to, the realization of the cognitive process or state of interest (Wright forthcoming). Techniques like subtraction and pattern classification analysis are not as important for addressing noise as they are for addressing the indirectness of the data. The difficulty of this task is a function of the knowledge investigators have about the causal links connecting the phenomena of interest and the measurements. In the case of the BOLD signal and neural activity, the fine-grained details of those causal connections are mostly unknown.

Data analysis techniques — such as subtraction or pattern classification analysis — are used because they isolate patterns in the data that are informative about the claims,

hypotheses or phenomena under investigation. New data analysis techniques change the evidential value of data by making it possible to detect phenomena via the data patterns they isolate. They do not do so by directly detecting the influence of the phenomena. If they were supposed to detect the influence of phenomena directly, then the skeptical arguments that point out worrying assumptions (e.g., van Orden and Paap 1997; Ritchie, Kaplan and Klein forthcoming), would be sufficient to undermine the capacity of data patterns to aid in assessment of the evidential value of neuroimaging data. It is here that a distinction between data and the data patterns is informative. Such a distinction, like the distinction between data and phenomena, can be drawn by appeal to the different evidential roles data and data patterns play in neuroimaging research.

Data are often recognized by philosophers of science as valuable as evidence for claims about phenomena to the degree that it is produced in such a way that the phenomenon of interest exerts a detectable influence on it. Bogen and Woodward notice that data have a number of undesirable features, in that they are complex and idiosyncratic to the context in which they are produced. These features are tolerated because they allow "… data to be useful as evidence…" (1988, p. 319). Ian Hacking recognized data as the 'marks', produced by interactions between experimenters, measuring instruments and the objects of measurement (1992). The most recent account of data made available in the philosophical literature goes a step beyond this, arguing that data are anything that are used as evidence (Leonelli 2015). This places the need to use data as evidence as more than just a constraint on data production, but constitutive of what data are.

Experimental practices, such as carefully controlling experimental environments and the efforts scientists go through to fix and stabilize observations, are often identified as the central facilitators of the 'detectability' of the influence of the target phenomena. This is part of Woodward's argument that 'building a better microscope' is the most common path to better science (2000), and these practices are the source of complexity and idiosyncrasy that Bogen and Woodward identify as a necessary for the pursuit of experimental knowledge (1988). The preceding arguments and examples I have presented make clear that, as is the case in neuroimaging research, when experimental controls are unable to ensure that there is a clear and clean causal link between the objects of

measurement and the target phenomenon, data analysis and manipulation practices are used to ‘cross the gap’. They do this by identifying, or isolating, patterns in data that could possibly reflect the influence of the target phenomenon. Data patterns, however, do not provide sufficient evidence to warrant claims about that phenomenon.

As I argue in the third chapter, data patterns are ill suited as evidence for claims about phenomena because they are the product of manipulations that suppress information relevant for evaluating claims about phenomena. This is consistent with the skeptical arguments that identify the decisions involved in data analysis as a source of inferential error (such as Aktunç 2014; Mole and Klein 2010). Skeptical arguments criticizing inferences in neuroimaging research by focussing on the limitations of data analysis techniques identify genuine inferential limitations of data patterns, but then mistakenly ascribe them to the data the patterns are derived from. This move conflates the evidential value of data with the value of data patterns. Data patterns are evidence for claims pertaining to the evidential relation that holds between data and phenomena, and not evidence for the claim about phenomena directly.

If data patterns are used to evaluate the evidential relation that holds between data and claims about phenomena, then what criteria do, and ought to, guide their isolation and explanation? In the fourth chapter, I argued that the choice of data analysis techniques, and decisions made during their application, are guided by investigator’s understanding of the function of the technique, the background theory they are approaching the data from the perspective of, and the research questions they are analyzing it to answer. In chapter three, I argued that data patterns are valuable insofar as they can be explained by appeal to claims about phenomena. Putting these together, a data pattern is explicable within a context when the theoretical background and understanding of the analysis technique used to isolate it provides the explanatory resources to connect claims about the target phenomena to the data pattern. Changes in any of these factors — analysis techniques, theoretical background or understanding of the analysis process — can change the perceived value of data by altering the explanatory relations investigators perceive between claims about phenomena and data patterns.

In summary, for data to be evidence for a claim about a phenomenon the causal factors that give rise to the phenomenon must (1) be involved in the production of the data such that they (2) leave detectable patterns in the data. The first condition reflects the fact that data cannot provide evidence for a phenomenon that played no role in its production. The second constraint is epistemic. Data cannot provide evidence for a phenomenon that played a role in its production, but leave no detectable patterns in the data. Data production processes are important for ensuring that the first condition is satisfied, while the second condition is contingent on the methods and processes used to isolate and interpret patterns in the data. Data patterns are valuable because they allow researchers to assess the relevance and significance of data with respect to a claim about phenomena. Data patterns are not a panacea to the challenges of interpreting data and determining its evidential value. They, and the analysis techniques used to isolate them, are better regarded as tools that play a central role in engaging and overcoming those challenges.

I argued in the preceding papers that the inferential gap between data and claims about phenomena is managed by the use of multiple analysis techniques to isolate and explain multiple patterns. In the next section I return to this idea and re-examine the significance of multiple data patterns given the challenges with using neuroimaging data as evidence outlined above. Where I had suggested in the second chapter that they provide inferences with a degree of robustness, taking into account the inferential limitations of data patterns and the process by which they are created as discussed above, robustness doesn't quite fit as an account of the epistemic advantage provided by multiple data patterns.

5.3 Multiple Patterns and Robustness

Interpreting neuroimaging data involves using a variety of analysis techniques. Each technique produces a different data pattern by imposing different manipulations and transformations on the data. Each manipulation, if the resulting data pattern is to be interpreted as informative about the target phenomena, involves making assumptions about the data. This is what skeptics often pick up on — arguing that subtractive methods assume that discrete regions have discrete function (van Orden and Paap 1997), and that classification analysis assumes successful classification is indicative of informational content (Ritchie, Kaplan and Klein forthcoming). The example I considered in the second

chapter showed how analysis procedures involving distinct assumptions can be used to limit the reliance on assumptions in the final inference. In particular, the use of a permutation test to select voxels for the pattern classification analysis minimized the extent to which the inference relied on the assumption that the classifier was leveraging relevant patterns in the BOLD data. I concluded the paper by drawing a parallel with discussions of robustness in the context of debates about the epistemic advantages of the use of multiple modes of measurement to validate measurement devices. When the same result is obtained by multiple, independent, processes that result is regarded as robust (Wimsatt 1981, p. 61). The suggestion was that multiple analysis processes enhance inferences in neuroimaging research by providing a more robust body of evidence than the results of a single analysis technique. Robustness is desirable because each line of supporting evidence is independent, ensuring that the result stands even if some of the evidence for it is overturned.

While this use of multiple analysis techniques in the interpretation of neuroimaging data appears to fit the model of robustness, I did cautiously classify the robustness of a collection of data patterns as weak and local. These qualifications are noteworthy because, upon closer inspection, they each cut against one of the core features of robustness. A conclusion is robust when multiple independent lines of evidence converge on it. A collection of data patterns is weak because the outcomes are not directly comparable and local because they are derived from a shared data set. With the arguments and examples from previous chapters now in hand, as well as the distinction between data and data patterns, it is worth looking again at these qualifications. I consider the locality and weakness each in turn.

5.3.1 Local and Dependent

The dispute examined in the fourth chapter, over the reliability and significance of NeuroSynth data for claims about the selectivity of discrete regions of the brain, shows that the significance of data is determined by locally shaped epistemic criteria. The criteria investigators consider when making judgements of data's relevance are not passively applied to the results of analysis techniques, but actively direct decisions made during the analysis process. In this way, theoretical goals and the conceptual

understanding of data analysis procedures that determine the epistemic criteria data patterns are evaluated by also influence decisions made in the process of creating those patterns. While the concerns about locality raised in the second chapter mostly had to do with the data patterns originating from the same data set, the case examined in chapter four suggests that the shared theoretical context may be a bigger threat to their independence. Alison Wylie's work on inferences in archaeology puts a finer point on the problem here.

Wylie distinguishes two dimensions along which research practices can be independent: theoretical independence, which concerns the background theories, auxiliary hypotheses and modes of reasoning involved in the practice; and causal independence, which concerns the causal factors that give rise to a given data set or body of evidence (1999, p. 304). She argues that both are necessary for convergent results to be truly robust, and cautions, as many contributors to the work on robustness analysis do, that appearances of independence can lead to mistaken confidence in claims inferred on their basis (also see Wimsatt 1981; Calcott 2013).

The lines of evidence represented by a distinct analysis process are conceptually dependent. They are produced within the same research context and decisions made along the way are guided by the same theoretical goals and background. While there is a limited degree of causal independence — different patterns emphasize the influence of some causal factors, suppress the influence of others, and often require different assumptions of the data to obtain for their results to be interpretable — the patterns ultimately originate from the same data set. This creates the conditions that can lead to what Wimsatt calls illusions of robustness (1981, p. 71). Illusions of robustness can occur when the appearance of independence conceals the dimensions along which the results and methods are the same. The dispute over the value of NeuroSynth data provides a compelling example of how the failure of the process by which data patterns are arrived at to be theoretically independent can lead investigators to be mistakenly confident in a result.

Lieberman and Eisenberger's mistaken assessment of the data's value is driven by the theoretical perspective they approach data interpretation from. Their theoretical perspective, informed by their prior work on dACC function and understanding of the meaning of reverse inference maps, directs the decisions they make in the analysis of the data. In particular, the decision to only analyze reverse inference maps for a small selection of terms, and discounting the significance of posterior probabilities. Both decisions are justified by appeal to their theoretical commitments (Lieberman 2015), and together they lead them to regard the data as warranting evidence for a selectivity claim. L&E present multiple lines of evidence in the form of reverse inference maps to support this conclusion. A conclusion ultimately dependent on the theoretical perspective they approached the analysis from, as Yarkoni's criticism — which is made from a distinct theoretical perspective — demonstrates.[24]

---

[24] Recall that Lieberman, in response to Yarkoni's argument that the posterior probability estimates are the right data pattern to consider when evaluating the significance of NeuroSynth data for the target claim, insisted that they were 'not interested' in posterior probabilities. This reflects a judgement that a data pattern, which Yarkoni presents as contradictory evidence for the claim that dACC is pain selective, is actually irrelevant.

This example echoes a central premise of Jacob Stegenga's argument against the significance of robustness (2009). According to Stegenga, it is rare for a diverse body of evidence to be convergent, and in cases where there is discordance the pursuit of robustness does not help because the independence of the multiple modes of evidence is often the source of the problem (p. 658). Indeed, the dispute over the value of NeuroSynth data appears intractable in part because of the theoretical differences which lead to opposing judgements of evidential value. Yarkoni's lack of expertise with respect to research on dACC function, and Lieberman and Eisenberger's lack of expertise with respect to the curatorial procedures of NeuroSynth, are referred to in arguments against the validity, or relevance, of the opposing interpretation. Stengenga notes that decisions about the relevance of data may be able to resolve the problem of discordant evidence.

Data patterns isolated by multiple analysis techniques applied to a neuroimaging data set are both causally and theoretically dependent. The shared theoretical background of a collection of data patterns is a greater inferential risk than the shared origins. Not only can the theoretical background data patterns be considered within direct assessments of evidential value, but it also guides the process of creating those data patterns and can direct judgements of the relevance of contradictory results provided by independent analyses. Furthermore, the data interpretation process is aimed at determining the evidential value of a particular data set and so, if multiple analysis processes are to be used, it only makes sense to apply them to that data set. Theoretical dependence is an epistemic liability, and causal dependence is unavoidable. Next, I consider the relationship between the 'weakness' qualifier and the requirement that a robust body of evidence converges on the same result.

5.3.2 Weak and Divergent

I classified the body of evidence provided by multiple analysis techniques as weak because the patterns isolated by each technique are distinct and often cannot be directly compared. In a footnote, I remarked that, while this is a problem for evaluating the convergence of the results, which is required by a robust body of evidence, it isn't a problem unique to neuroimaging research. I referred to Jacqueline Sullivan's work on the multiplicity of experimental protocols, which casts doubt on the assumption that different experiments aiming at understanding the same phenomenon in fact instantiate the same phenomenon, given that the experimental protocols guiding the experiment often differ between research contexts (2009). An analogous problem holds for the use of multiple

---

However, this just shifts the problem of identifying criteria for adjudicating between discordant results, to identifying criteria for determining what results are relevant (p. 660). The same theoretical perspectives that lead to the opposed judgements of significance, also direct the arguments against the relevance of the contradictory interpretation and its supporting data patterns.

data analysis techniques. Since they each implement a distinct series of manipulations, isolating distinct patterns by suppressing different facets of the data, their results are not about 'the same' thing.

Consider how the results of univariate techniques and pattern classification analysis are used together in practice (as discussed in Coutanche 2013; Davis et al 2014). Subtraction techniques proceed by averaging BOLD signal data within a region of interest then contrasting that regional average between two task conditions. The result indicates the mean difference in activity between the task conditions, or the mean activation of the region. Pattern classification analysis involves using one set of variables, usually BOLD activity, to predict another, either task conditions or behavioural responses. Classifier accuracy is taken to indicate if information relevant to discriminating between the tasks is available in the signal. Marc Coutanche's discussion of the combined value of these tools indicates that some studies use univariate techniques, such as subtraction analysis, in series with pattern classification analysis in order to evaluate the sufficiency of the information carried by multivariate patterns, while others use them in parallel to evaluate the necessity of multivariate patterns for successful classification (2013, p. 669). The sequence approach involves using a univariate technique to remove the mean activation from the data and assess the effect on classifier performance. If the classifier accuracy remains high, then this is evidence that the multivariate pattern (that is, relative differences in activity between data points) is sufficient for classification. The parallel approach can be conducted in several ways, either by performing classification on the isolated mean activity, or by comparing subtraction and classification results. In either case, the aim is to determine if there is information available in one, both, or neither of the multivariate pattern and mean activity (p. 669-70). These comparisons would not be meaningful if the patterns isolated by subtraction and pattern classification analysis were convergent.

In characterizing the common features of concepts of robustness, Wimsatt notes that they often involve looking for and analyzing things that are "invariant over or identical in the conclusions or results of…" independent processes (1981, p. 44). The invariant results or conclusions are regarded as robustly supported, and are conferred additional support by

the independence of the various processes that produce them. In the comparison of mean activation and classifier accuracy, investigators are not trying to identify an invariant value, variable or pattern. Instead, claims about mean activation and claims about classifier accuracy are taken together to be informative about the how information may be encoded in the BOLD signal. If the mean activity is low, and the classifier is accurate with and without the mean levels, then investigators conclude that task relevant information is encoded in a multivariate pattern (that is, multiple parts acting in a coordinated way). Identifying invariant properties is not the aim of applying multiple analysis techniques to the data. Each technique is applied for the unique perspective it provides on the evidential significance of the data with respect to a specific set of claims and hypotheses.

This, and the failure of independence noted above, are together strong reasons to resist the robustness account of the value of multiple patterns. While multiple analysis results do not confer a data-phenomena inference with robustness, they are valuable for interpreting the data and that value follows from their distinctiveness, not their independence or convergence. William Bechtel, looking at how neuroscientists combine multiple research techniques to make inferences, provides an alternative to robustness along these lines. He argues that multiple research techniques are valued in neuroscience for their complementarity, not independence (2002). The same can be said of multiple data analysis techniques.

5.3.3 Complementary Perspectives

Techniques like single cell recording and neuroimaging are often used to calibrate one another, and so fail to be independent in the way required for a convergence of results to be regarded as an instance of robustness (Bechtel 2002, p. S49). Furthermore, techniques like single cell recording, involve invasive interventions, such as implanting electrodes into the brain, that can alter the functioning of the system. The same could be said about the manipulations involved in the production and analysis of neuroimaging data: they distort the data so that the data fail to reflect the full spectrum of causal factors that give rise to the phenomena of interest. In this way, these techniques "… provide a very selective and distorted perspective on the phenomena" that they are used to investigate (p.

S49). Bechtel argues that this is not necessarily a problem for inferences in neuroscience, as these differences can be, and are, leveraged to strengthen inferences and develop better theories. Each technique is able to answer specific questions about the relationship between brain activity and cognitive functioning. Taken together, they provide a more complete picture of the phenomena involved in the production of the data. For instance, where single cell recordings are limited to providing information about the function of specific pieces of the brain, neuroimaging can provide information about more widespread network-level activity (p. S54-5). The results of different techniques are complementary as each provides information about the target phenomena that the others cannot.

The same can be said of multiple analysis techniques as used to aid in the interpretation of neuroimaging data. Data analysis techniques transform data by suppressing some features and highlighting others in order to isolate a data pattern. The usefulness of a data pattern is its interpretability when compared to the often-complex data sets it is derived from. This is especially true in neuroimaging research, where data sets are large, and measurements are indirectly related to the phenomena neuroscientists are interested in learning about. Analysis techniques like subtraction and pattern classification analysis are used to pick out specific patterns in the data that are informative about some, but not all, aspects of the phenomena involved in its production. Data patterns do not reflect the full range of causal factors involved in the data's production that are relevant to the phenomena of interest. In other words, data patterns are selective distortions of the data.

Multiple data analysis techniques do not provide independent lines of evidence as much as they clarify the evidential import of the data with respect to claims about phenomena by virtue of the distinct, and distorted, perspectives they make available to investigators. Neuroimages and the machine learning classifier's accuracy at discriminating between conditions are each the product of a process that distorts the data. Each analysis result is an isolated data pattern that reveals specific features of the data set at the expense of being informative about other features. Distinct patterns warrant different claims about the data, which in turn can be explained by appeal to claims about phenomena. Their individual value is in their explicability, and their collective value in their distinctiveness.

While each analysis technique may require assumptions of the data that are potentially false, and the manipulative process itself invites a variety of alternative explanations for the pattern, the collection of multiple distinct patterns softens the impact of these complications. As argued above, data patterns are not used to infer claims about phenomena but are used to make claims about data and assess its value as evidence for claims about phenomena. The distinctiveness of a collection of data patterns provides a richer explanatory target than a single data pattern. Viable alternative explanations for a data pattern may be readily available when only one pattern is considered, but to remain viable the alternative must not be ruled out by another pattern in the data. The distinctiveness of the patterns improves the capacity of a large collection of them to rule out alternatives in this way, and their collective explanation provides resistance against alternatives not explicitly considered by the investigators.

This process has epistemic advantages, as discussed above and in the second and third chapter, but also brings with it risks, as demonstrated in the fourth chapter. Viewing the process of data interpretation as explanatory, and aimed at explaining data patterns, provides conceptual resources — and a perspective — that can aid in disentangling the various factors that contribute to inferences from data as complex, varied and difficult to make sense of as neuroimaging data. I have shown as much with this collection of work, but there is much more work to be done. Especially as the technologies and tools for handling, manipulating, analyzing and sharing neuroimaging data continue to rapidly evolve.

5.4 Looking Forward

I have focused, with the exception of chapter four, on the positive contribution data analysis makes to the interpretation of neuroimaging data. On the other side of the coin are the inferential errors and reasoning mistakes that these techniques and the explanatory process they contribute to, make possible. The increasing variety and complexity of analysis procedures can lead to inferential errors or the misinterpretation of data in a number of ways. Errors can be made in the data manipulations themselves, as a recent paper showing that a significant number of neuroimaging studies may be the product of systematic errors in standard analysis software and pipelines demonstrates (e.g., Eklund

et. al. 2016). Errors can also arise when investigators' understanding of the analysis technique misrepresents its actual function. More generally, some have raised a concern that the emphasis on greater sophistication in analysis may slow progress since a strict focus on data creates distance between researchers and the material objects and phenomena they are investigating. Familiarity with and close proximity to the material objects under study has historically been important for inspiring the ideas that have marked significant leaps of progress in the history of neuroscience (Marder 2015), and has been identified as necessary for the effective interpretation of data in general (Leonelli 2013). However, the work here shows that familiarity with the data analysis techniques is equally important for shaping intuitions informing judgements about the relevance of data patterns, the claims they warrant about the data, and explaining them by appeal to claims about phenomena. Concerns about analytic flexibility, such as those raised by Joshua Carp (2012), draw attention to inferential errors that can occur due to decisions made in the course of implementing a data analysis technique. These kinds of decisions have more impact when they pertain to pre-processing steps, which affect all subsequent analyses of the data, then if they occur in the interpretive stage where they only affect one data pattern. Analytic flexibility could also lead to prematurely ending data interpretation because the data patterns isolated first do not provoke alternative explanations, making the data appear to better support the target claim than it may have if other data patterns had been isolated.

All of this points to an epistemic tension in the development and uptake of novel techniques for analyzing data. On one hand, new techniques can clarify the evidential import of data with respect to competing claims and hypothesis and promote the discovery and study of phenomena previously impossible to detect in a data set. On the other, new analysis techniques have the potential to lead a field towards the ever more sophisticated production of misleading results. This situation is even more precarious in neuroimaging research, where the data sets investigators work with consist in a large number of variables — every six seconds a subject is in the scanner can result in over a thousand data points — and have relatively low power due to low participant counts, and limitations of experimental paradigms. Factor in the increasing variety of approaches to

data analysis, and there are sufficient grounds to argue that neuroimaging research is "… a 'perfect storm' of irreproducible results" (Poldrack et al 2017).

A common proposal for addressing this situation is to foster reproducible research practices. Reproducible practices are those which allow independent investigators to reconstruct the data analysis process using publicly accessible materials. Suggestions for doing so include sharing data, algorithms, analysis code, and the pre-registration of research plans (Poldrack et al 2017; Munafò et al 2017). The ideal reproducibility aims at is a practice in which independent investigators can reconstruct the data interpretation procedures that lead to the results of articles published by their peers. While the first step towards fully reproducible practices is to make data and the code used to analyze it accessible, achieving the ideal depends on the ability of investigators to identify, and contrast, the rationale for the decisions made at each step of the process. Disputes such as the case examined in chapter four show how doing so can be informative. Open discussion allows for errors in reasoning due to the misunderstanding of data analysis techniques, and mistaken judgements of data's relevance to a claim, to be unearthed and articulated. While it may not resolve the disagreement between the involved parties, the ability to reproduce the conceptual elements of the interpretive process is necessary for a practice to be fully reproducible. These factors permeate the interpretation process, and so may have a greater influence on the judgement of data's significance than the manipulations of the data themselves.

New tools and techniques for the analysis, handling, storing and classification of data are being developed and promoted in response to limitations of neuroimaging research methods, and the growing volume and complexity of both the data, and knowledge, produced by neuroscientific research. These tools include automatically curated databases that allow users to perform large-scale, automated meta-analyses, such as NeuroSynth (Yarkoni et al 2011), data repositories designed to handle the full complexity and diversity of neuroimaging data produced in experiments, such as OpenfMRI (Poldrack et al 2013; Poldrack and Gorgolewski 2015), and frameworks that support and facilitate the development of a community-driven knowledge base, or ontology, such as the Cognitive Atlas (Poldrack et al 2011). Critical analyses that examine how the interface between

data, theories, and communities, are changing and have been changed by these technologies will contribute to philosophical debates and could improve neuroscientific practice. The arguments of the preceding papers, the approach presented above, and a method of research that includes close interactions with scientists, provides a foundation for pursuing this investigation in a manner that could prove valuable for improving the inferential practice of those using these new techniques.

References

Aktunç, E. M. [2014]: 'Severe Tests in Neuroimaging: What We Can Learn and How We Can Learn It', Philosophy of Science, 81, pp. 961-73.

Anderson, M. [2015]: 'Mining the brain for a new taxonomy of the mind', Philosophy Compass, 10, pp. 68-77.

Anderson, M. L. and Oates, T. [2010]: 'A critique of multi-voxel pattern analysis', Proceedings of the 32nd Annual Conference of the Cognitive Science Society, pp. 1511-16

Bogen, J., and Woodward J. [1988]: 'Saving the Phenomena', Philosophical Review, 97, pp. 303-52.

Bechtel, W. [2002]: 'Aligning multiple research techniques in cognitive neuroscience: Why is it important?', Philosophy of Science, 69, pp. S48-S58.

Calcott, B. [2011]: 'Wimsatt and the robustness family: Review of Wimsatt's Re-engineering Philosophy for Limited Beings', Biology and Philosophy, 26, pp. 281-93.

Carp, J. [2012]: 'On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments', Frontiers in Neuroscience, 6, 149.

Coutanche, M. N. [2013]: 'Distinguishing multi-voxel patterns and mean activation: Why, how, and what does it tell us?', Cognitive Affective Behavioural Neuroscience, 13, 667-673.

Davis, T., LaRocque, K. F., Mumford, J. A., Norman, K. A., Wagner, A. D., and Poldrack, R. A. [2014]: 'What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis', Neuroimage, 97, pp. 271-83.

de-Wit, L. Alexander, D., Ekroll, V., and Wagemans, J. [2015]: 'Is neuroimaging measuring information in the brain?', Psychonomic Bulletin and Review, 23, pp. 1415-1428.

Eklund, A., Nichols, T. E., and Knutsson, H. [2016]: 'Cluster failure: Why fMRI inferneces of spatial extent have inflated false-positive rates', Proceedings of the National Academy of Sciences of the United States of America, 113(28), pp. 7900-7905.

Hacking, I. [1992]: ‘The self-vindication of the laboratory sciences’, in A. Pickering (ed), Science as Practice and Culture. University of Chicago press, pp. 29-64.

Haxby, J. V. [2010]: ‘Multivariate Pattern Analysis of fMRI data’ in M. Bunzl and S. J. Hanson (eds), Foundational Issues in Human Brain Mapping, The MIT Press, pp. 55-68.

Kohler, P.J., Fogelson, S.V., Reavis, E.A., Meng, M., Guntupalli, J.S., Hanke, M., Halchenko, Y.O., Connolly, A.C., Haxby, J.V. and Peter, U.T., [2013]: ‘Pattern classification precedes region-average hemodynamic response in early visual cortex’. NeuroImage, 78, pp. 249-260.

Landreth, A. and Richardson, R. C. [2004]: ‘Localization and the new phrenology: A review essay on William Uttal’s the new phrenology’, Philosophical Psychology, 17, pp. 107-23.

Lenartowicz, A., Kalar, D. J., Congdon, E., & Poldrack, R. A. [2010]: ‘Towards an ontology of cognitive control’, Topics in Cognitive Science, 2(4), pp. 678-692.

Leonelli, S. [2015]: ‘What Counts as Scientific Data?’, Philosophy of Science, 82(5), pp. 810-821.

Leonelli, S. [2013]: ‘Data Interpretation in the Digital Age’, Perspectives on Science, 22, pp. 397-417.

Lieberman, M. D. [2015]: ‘Comparing Pain, Cognitive, and Salience accounts of dACC’. Available at: https://www.psychologytoday.com/blog/social-brain-social-mind/201512/comparing-pain-cognitive-and-salience-accounts-dacc

Lindquist, M. A., Loh, J. M., Atlas, L. Y., and Wager, T. D. [2009]: “Modelling the hemodynamic response function in fMRI: Efficiency, bias and mis-modelling”. NeuroImage 45(1): p. S187-S198.

Machery, E. [2014]: ‘Significance Testing in Neuroimagery’, in Kallestrup J., and Sprevak, M.(eds.), New Waves in the Philosophy of Mind, Palgrave Macmillan, pp. 262-277.

Marder, E. [2015]: ‘Understanding Brains: Details, Intuitions, and Big Data’, PLoS Biology, 13(5), e1002147.

Martin, C. B., Cowell, R. A., Gribble, P. L., Wright, J., Köhler, S. [2015]: ‘Distributed category-specific recognition memory signals in human perirhinal cortex’, Hippocampus.

Mole, C. and Klein, C. [2010]: ‘Confirmation, Refutation, and the Evidence of fMRI’, In Stephen Hanson & Martin Bunzl (eds.), Foundational Issues in Human Brain Mapping. Cambridge: MIT Press. pp. 99-112.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E., Ware, J. J., and Ioannidis, J. P. A. [2017]: ‘A manifesto for reproducible science’, Nature Human Behaviour, 1.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò M. R., Nichols, T. E., Poline, J., Vul, E., and Yarkoni, T. [2017]: ‘Scanning the

horizon: towards transparent and reproducible neuroimaging research', Nature Reviews Neuroscience, 18, pp. 115-126.

Poldrack, R. A., and Gorgolewski, K. J. [2015]: 'OpenfMRI: Open sharing of task fMRI data', NeuroImage, 144, pp. 259-261.

Poldrack, R. A., Barch, D. M., Mitchell, J. P., Wager, T. D., Wagner, A. D., Devlin, J. T., Cumba, C., Koyejo, O., and Milham, M. P. [2013]: 'Towards open sharing of task-based fMRI data: the OpenfMRI project', Frontiers in Neuroinformatics, 7(12). doi:10.3389/fninf.2013.00012

Poldrack, R.A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D. S., Sabb, F. W., and Bilder, R. M. [2011]: 'The Cognitive Atlas: Towards a Knowledge Foundation for Cognitive Neuroscience', Frontiers in Neuroinformatics, 5(17). doi:10.3389/fninf.2011.00017

Ritchie, J.B., Kaplan, D.M., and Klein, C. [Forthcoming]: 'Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience', British Journal for the Philosophy of Science.

Roskies, A. [2010]: 'Saving Subtraction: A reply to Van Orden and Paap', British Journal for the Philosophy of Science, 61, pp. 635-65.

Stegenga, J. [2009]: "Robustness, Discordance, and Relevance", Philosophy of Science, 76, pp. 650-661.

Sullivan, J. [2009]: 'The multiplicity of experimental protocols: A challenge to reductionist and non-reductionist models of the unity of neuroscience', Synthese, 167, pp. 511-39.

Tambini, A., and Davachi, L. [2013]: 'Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory', Proceedings of the National Academy of Sciences of the United States of America, 110(48), pp. 19591-19596.

Tong, F., and Pratte, M. S. [2012]: 'Decoding Patterns of Human Brain Activity', Annual Review of Psychology, 63, pp. 483-509.

van Orden, G. C., and Paap, K. R. [1997]: 'Functional Neuroimages Fail to Discover Pieces of Mind in Parts of the Brain', Philosophy of Science, 64, pp. S85-94.

Wimsatt, W. [1981]: 'Robustness, Reliability, and Overdetermination", in Re-Engineering Philosophy for Limited Beings, pp. 43-74.

Woodward J. [2000]: 'Data, phenomena, and reliability', Philosophy of Science, 67(3), pp. S163-S179.

Wright, J. W. [Forthcoming]: 'The Analysis of Data and the Evidential Scope of Neuroimaging Results', British Journal for the Philosophy of Science.

Wylie, A. [1999]: "Rethinking Unity as a "Working Hypothesis" for Philosophy of Science: How Archaeologists Exploit the Disunities of Science", Perspectives on Science, 7, pp. 293-317.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. [2011]: 'Large-scale automated synthesis of human functional neuroimaging data', Nature Methods, 8, pp. 665-670.

Curriculum Vitae

Name: Jessey Wright

Post-secondary Education and Degrees:

The University of Waterloo
Waterloo, Ontario, Canada
2006-2011 B.Math.

The University of Waterloo
Waterloo, Ontario, Canada
2011-2012 M.A.

The University of Western Ontario
London, Ontario, Canada
2012-2017 Ph.D.

Honours and Awards:

Social Science and Humanities Research Council (SSHRC)
Doctoral Fellowship
2013-2016

Michael Smith Foreign Study Supplement (via SSHRC)
2015

Rotman Institute of Philosophy Catalyst Funding
'The Epistemic Status of fMRI Technology and Data Analysis Techniques'
with Köhler, S., Sullivan, J., Martin, C.
2014

Ontario Graduate Scholarship
2012

Related Work Experience

Research Assistant
Rotman Institute of Philosophy
Lab Associates Curriculum Project
with Weijer, C., Sullivan, J., and Foley, R.
2015

Related Publications:

Wright, J. (Forthcoming). The Analysis of Data and the Evidential Scope of Neuroimaging Results. British Journal for the Phliosophy of Science.

Martin, C.B., Sullivan, J. A., Wright, J., and Köhler, S. (Forthcoming). Recognition-memory signals for objects from different categories are graded across perirhinal and parahippocampal cortex. NeuroImage.

Martin, C.B., Cowell, R.A., Gribble, P.L., Wright, J., and Köhler, S. (2015). Distributed category-specific recognition memory signals in human perirhinal cortext. Hippocampus. DOI: 10.1002/hipo.22531