Prospects for a Data-Driven Task-fMRI Ontology

Taylor Sullivan Bolt
University of Miami, tsb46@miami.edu

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation
https://scholarlyrepository.miami.edu/oa_dissertations/2066

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.
UNIVERSITY OF MIAMI

PROSPECTS FOR A DATA-DRIVEN TASK-FMRI ONTOLOGY

By

Taylor Sullivan Bolt

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

May 2018
UNIVERSITY OF MIAMI

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

PROSPECTS FOR A DATA-DRIVEN TASK-FMRI
ONTOLOGY

Taylor Sullivan Bolt

Approved:

__________________________  ________________________
Lucina Q. Uddin, Ph.D.       Maria M. Llabre, Ph.D.
Associate Professor of Psychology  Professor of Psychology

__________________________  ________________________
Aaron S. Heller, Ph.D.       Roger C. McIntosh, Ph.D.
Assistant Professor of Psychology  Assistant Professor of Psychology

__________________________  ________________________
Michael L. Anderson, Ph.D.   Guillermo Prado, Ph.D.
Professor of Philosophy      Dean of Graduate School
Western University
The growing literature of task-based functional magnetic resonance imaging (task-fMRI) has increased calls for an adequate organizing ontology, or taxonomy, of task fMRI experiments. Researchers differ over what should be the dominant features of such an ontology: should it be concrete/observable dimensions, such as task paradigm (e.g. n-back vs. flanker task), or latent/theoretical dimensions, such as cognitive domains (e.g. working-memory vs. bottom-up attention)? This dissertation attempts to address what is important for a task-fMRI ontology in a quantifiable manner. We use a simple quantitative criterion for categories/dimensions of a task-fMRI ontology: the ability to explain observed variations in task-fMRI activation patterns. Using meta-analysis tools and multivariate statistical methods, we identify those dimensions and categories of the task-fMRI environment that explain observed variations in task-fMRI activation patterns. In study one, we observed that a preliminary ontology of four or seven latent cognitive categories provides a simplified description of the observed differences in whole-brain blood-oxygen-level dependent (BOLD) activation patterns. In study two, we observed that while these categories may provide an adequate description of BOLD activation patterns at the population level, inter-subject variability restrains inferences from these population level distinctions to the subject level. In conclusion, results from both studies suggest that a data-driven task-fMRI ontology is a viable project for cognitive neuroscience.
# TABLE OF CONTENTS

**LIST OF FIGURES**

---

**Chapter**

1 INTRODUCTION
   - The Need for a Data-Driven Task-fMRI Ontology ........................................ 1
   - Two Kinds of Task-fMRI Ontologies .............................................................. 3
   - Concrete and Latent Dimensions of a Task-fMRI Ontology ............................. 7
   - Strategies for Building a Task-fMRI Ontology .............................................. 15
   - Analysis Framework for the Dissertation ..................................................... 23

2 STUDY ONE
   - Background .......................................................................................................... 28
   - Methods .................................................................................................................. 29
   - Results .................................................................................................................. 41
   - Discussion ............................................................................................................. 52

3 STUDY TWO
   - Background .......................................................................................................... 58
   - Methods .................................................................................................................. 60
   - Results .................................................................................................................. 66
   - Discussion ............................................................................................................. 72

4 SUMMARY ............................................................................................................... 77

**WORKS CITED** .................................................................................................. 78

**APPENDIX** ....................................................................................................... 88
LIST OF FIGURES

Figure 1 ........................................................................................................... 20
Figure 2 ........................................................................................................... 22
Figure 3 ........................................................................................................... 32
Figure 4 ........................................................................................................... 42
Figure 5 ........................................................................................................... 45
Figure 6 ........................................................................................................... 47
Figure 7 ........................................................................................................... 50
Figure 8 ........................................................................................................... 51
Figure 9 ........................................................................................................... 63
Figure 10 ........................................................................................................ 68
Figure 11 ........................................................................................................ 70
Figure 12 ........................................................................................................ 70
Figure 13 ........................................................................................................ 72
Chapter 1: Introduction

The Need for a Data-Driven Task-fMRI Ontology

Imagine a researcher using functional magnetic resonance imaging (fMRI) to discover the brain areas that activate to a simple inhibition task (e.g. Eriksen Flanker Task). According to the standard experimental design used in cognitive neuroscience, the researcher would attempt to isolate the supposed “inhibition process” with a carefully designed control and active condition. The control condition would match the active condition in every way possible outside of the “inhibition process” of interest, which might include matching the stimulus modality (e.g. visual), the motor response (e.g. button press), the visual stimuli (e.g. arrows or words), or the task demands (e.g. attend and answer). Given a sufficiently well-designed experiment, the researcher might be able to isolate those brain regions that selectively activate to “inhibition”.

Implicit in the researcher’s mind are the features of the task-fMRI experiment that matter most to the conclusions they draw from their results. In other words, the researcher knows those features of the task-fMRI experiment that need to be controlled for, and those that are superfluous. For example, the researcher obviously does not think that the participant’s eye color needs to be controlled for. Thus, eye color would be free to vary outside of experimental control. However, the researcher would be concerned with controlling for participant handedness, given the well-known differences in neural organization between left- and right-handed participants (Knecht et al., 2000). Other factors worth controlling for might be IQ, gender, or psychiatric diagnosis.
The point of the above discussion is that there are potentially infinite many ways to describe a task-fMRI experiment, or any experiment for that matter. Importantly though, some of the distinctions we might make are worth pointing out and considering in our design of the task-fMRI experiment. The reason we think these factors are important for task-fMRI is because of the consequences they may have on the neural activity, or more properly - blood-oxygen-level dependent (BOLD) activity, we observe in our experiment. If we assume that handedness drives differences between BOLD activation maps obtained from different subjects, we need to take this into account. In other words, we need to measure handedness, communicate it in our scientific results, and if it is not of interest in our study, experimentally or statistically control for it.

These concerns center around an ontology or taxonomy for task-fMRI. An ontology for a scientific field identifies the relationships and entities of their domain of study and describes them in a standardized vocabulary. In other words, it provides the categories or distinctions worth making and communicating to other researchers in their domain of study. Thus, an ontology based on task-fMRI data would categorize and communicate those features of the task-fMRI experiment that are important for researchers. A crucial factor in a task-fMRI ontology project is addressing what exactly is important. Researchers could reasonably disagree over whether some distinctions are worth making and communicating to fellow researchers. For example, all task-fMRI researchers would presumably not think it important to label participant eye color as an important feature of the task-fMRI experiment that needs codifying in a taxonomy. However, researchers could possibly disagree over whether the time of the day of data
collection is an important feature of the task-fMRI experiment worth controlling for and communicating to other researchers.

This dissertation attempts to address what is important for a task-fMRI ontology in a quantifiable manner. To anticipate the central argument, I believe the quality of such an ontology is how well it describes differences in task-fMRI activation patterns. Distinguishing between category/process A and category/process B in the task-fMRI environment when differing them produces no change in BOLD activation patterns does not contribute meaningfully to a task-fMRI ontology. In other words, this is a distinction without a difference. To be clear, this doesn’t mean there isn’t a difference, as I discuss below. But I do argue that it may be an unnecessary distinction to make in task-fMRI. Using this conception of a task-fMRI ontology project, we use meta-analysis tools and multivariate analysis techniques to answer these questions in a quantifiable manner.

Two Kinds of Task-fMRI Ontologies

A scientific ontology can refer to more than one thing: 1) a standardized terminology of entities and relationships that describe a scientific domain, or 2) the specification of what entities or categories exist in a scientific domain. The former definition describes work in the field of informatics or information science (Poldrack et al., 2011; Shepherd et al., 1998; Turner & Laird, 2012). Informatics is a research field concerned with the organization, dissemination and classification of data. The latter definition of ontology is more commonly used in the philosophy of science (Anderson, 2015; Janssen, Klein, & Slors, 2017; Price & Friston, 2005). This definition concerns the more theoretical concern of whether the terminology we utilize in our description of our
scientific domain refer to “natural” divisions of that domain. In other words, are these
categories or divisions “natural kinds”, or groupings of nature (Ellis, 2001; Hacking,
2000)? To be sure, these two definitions of ontology are connected in important ways,
which we discuss below. In the next two sections, we describe these two definitions of
ontology and how they map on to the task-fMRI ontology project.

Neuroinformatics and Standardized Vocabulary

Neuroinformatics has been developed in response to the increasing size and
complexity of neuroscience data. The accumulation of large amounts of complex data is
now a fact of life for many current scientific fields (Khoury & Ioannidis, 2014; Lynch,
2008; Marx, 2013). This “Big Data” revolution is central for neuroscience as well,
particularly neuroimaging (Lichtman, Pfister, & Shavit, 2014; Poldrack & Gorgolewski,
2014; Sejnowski, Churchland, & Movshon, 2014). Inspired by informatics success in the
gene-sequencing community during the 1980s, the field of neuroinformatics hopes to
provide informatics tools to the neuroscience community to more effectively
communicate results using advanced automated databases, computational search tools,
and standardized vocabularies (Shepherd et al., 1998). Neuroinformatics aims to supplant
the more traditional mode of neuroscientific research, where data is collected by single
laboratory, a small portion of that data is subsequently published, and the laboratory then
moves on. In the age of big data neuroscience, neuroinformatics promises to increase the
availability and utility of large-scale neuroscientific datasets.

However, this often-cited conception of neuroinformatics as a tool for organizing
and sharing large complex data misses another important function of neuroinformatics.
An important component of neuroinformatics is an *ontology* or standardized terminology for the description and sharing of neuroscience data, regardless of whether this data is “big” or “small”. A good ontology provides a concise and standardized terminology for communication between researchers. This standardized terminology is not just for sharing results in databases, but for *describing* them, regardless of whether these results are shared. An example of this latter component of neuroinformatics is Russel Poldrack’s Cognitive Atlas (CA) Project for cognitive neuroscience (Poldrack et al., 2011). The CA project was envisioned as a standardized database of concepts within cognitive science, including mental tasks and mental processes. The goal of the CA project is to provide both a clearer technical vocabulary and a well-annotated database for theory testing and meta-analysis in cognitive neuroscience.

Related to the CA project, is the cognitive paradigm ontology (CogPO) developed by Turner and Laird (Turner & Laird, 2012). CogPO was developed out of the need to characterize and classify the many cognitive/behavioral task paradigms in the cognitive science literature. There are a notoriously large amount of cognitive/behavioral task paradigms used in the cognitive science literature. Using the terminology of the BrainMap database (P. T. Fox & Lancaster, 2002), the goal of CogPO is to provide a standardized structure for describing cognitive/behavioral tasks. Similar to the CA project, the hope of CogPO is to provide an easier way to communicate results for theory testing and meta-analysis in cognitive science.

Both the CA and CogPO ontologies hold promise as standardized descriptive frameworks for an overall task-fMRI ontology. For example, given a sample task-fMRI experimental design, a researcher can communicate the “cognitive process” studied using
their task, and describe the details of the task paradigm (stimulus modality, response type, etc.) to the larger cognitive science community. However, a crucial question left unanswered in these ontologies is the relationship of these descriptive categories to the empirical domain they describe. This concerns the “natural kinds” sense of ontology.

**Task-fMRI Ontology of Natural Kinds**

This dissertation is primarily concerned with the “natural kinds” sense of ontology. A task-fMRI ontology in this sense concerns the categorization or taxonomy of the natural features of the task-fMRI experiment, meaning those features that the activity recorded using fMRI is sensitive to. This might include categories for stimulus type, such as words, faces, houses, or fixation crosses. It might include categories for response type, such as button presses, oral/facial movements, eye blinks, or speaking. It also might include categories for task paradigm, such as Eriksen Flanker Task, Stroop Task, n-back task, or delayed match to sample task. More controversial features of a task-fMRI ontology may be cognitive/behavioral categories, such as task-switching, inhibition, response selection, or fear processing, which we discuss below.

Importantly, the “natural kinds” conception of a task-fMRI ontology is deeply interconnected with the neuroinformatics one. The usefulness of a shared, standardized vocabulary for task-fMRI researchers is grounded in its ability to not only communicate results in a standardized vocabulary, but to communicate results in a natural vocabulary. By natural, I mean categories or divisions of the task-fMRI experiment that respect the divisions of the brain we observe using task-fMRI. For example, an ontology of protein classes would be less effective were it to make divisions between proteins that belong to
the same class, regardless of whether this presents a standardized vocabulary for communication. Thus, the effectiveness of a task-fMRI ontology for cognitive neuroscience is crucially dependent on whether its divisions or categories are respected by observation of BOLD activation patterns elicited by experimental tasks completed in the fMRI environment.

**Concrete and Latent Dimensions of a Task-fMRI Ontology**

Up to now, no distinction has been made between those features of a task-fMRI ontology that might be described as concrete or observable features, and those described as latent features. Concrete features of the task-fMRI experiment are observable features that are directly accessible to the researcher. Major concrete features often reported in task-fMRI experiments are stimulus modality (auditory vs. visual stimulus presentation), response types (button press, foot movement, speaking), and task paradigm (n-back, affective pictures, film viewing). Concrete features such as stimulus modality are often considered crucial features of the task-fMRI experiment, and if not of interest, should be adequately controlled for. For example, most task-fMRI researchers would argue that a contrast between two conditions to isolate “inhibition” would be confounded if the stimulus modality were different between the two conditions (e.g. auditory vs. visual). Thus, stimulus modality would presumably have a central place in a task-fMRI ontology. However, not all concrete features would be considered to have an important role in a task-fMRI ontology. For example, most researchers would agree that the weather outside of the MRI scanner room is not relevant in a task-fMRI ontology. Other less important concrete features of the task-fMRI scanner would be the shape of the MRI scanner, the
color of the experimenter’s shirt, or whether the experiment was conducted in April or June (and the list could go on). Now these are purposefully ridiculous examples that would obviously have no use in a task-fMRI ontology. However, factors such as caffeine intake, peripheral nerve pain, history of tobacco use, etc., represent features of the task-fMRI experiment that experimenters may disagree over the importance of. Thus, even observable, concrete features of the task-fMRI experiment require empirical assessment of their contribution to a task-fMRI ontology. But the disagreement over the importance of concrete features in the task-fMRI experiment pale in comparison to disagreement over latent features.

*Latent Cognitive Processes*

In cognitive neuroscience, differences in whole-brain BOLD activity between two conditions are generally inferred to represent differences in cognitive processes (Berman, Jonides, & Nee, 2006; Kanwisher, 2000; Poldrack, 2006; E. E. Smith, Jonides, & Koepppe, 1996; Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). With all due respect to behaviorists, most task-fMRI researchers agree that cognitive process terms refer to latent constructs. In other words, they are not directly observable, but inferred by observation. Although cognitive processes are the dominant latent feature of interest for most task-fMRI researchers, other latent features, such as mental representations are often studied. For example, in the context of multivariate pattern analysis (MVPA) differences in patterns of BOLD activity within a brain region are often inferred to constitute different mental representations in that region (Haxby, 2012).
While the exact definition of a cognitive process will differ between task-fMRI researchers, the vast majority of extant “process” terms in cognitive science are grounded explicitly or implicitly in the dominant information-processing account of human cognition. So powerful is this account in cognitive psychology that it is often considered synonymous with cognitive psychology (Baddeley, 2010; Neisser, 1967). According to this account, cognitive operations are understood by the functional role they play in an information-processing system as it transforms sensory representations from the environment into motor outputs. The cognitive system is often described in analogy to a digital computer. In a manner analogous to a computer, the mind receives and manipulates external or internal symbols and produces outputs. In the same manner that box-and-arrow plots are used to understand the serial operations of computers, box-and-arrow plots are used to understand the serial operations of the mind. The ubiquity of box-and-arrow plots in cognitive psychology and neuroscience publications testifies to the dominance of this conception of cognitive processes in both fields (Gazzaniga, Ivry, & Mangun, 2013; Korb, Wood, Rychlowska, & Niedenthal, 2016; Neisser, 1967), but see (Bolt, Uddin, & Anderson, In Press).

*A Brief Survey of the Cognitive Ontology Debate*

There is general agreement that some kind of ontology of latent cognitive processes is needed in an overall task-fMRI ontology, but whether the existing ontology of cognitive processes is sufficient is a matter of much serious debate (Anderson, 2015; Klein, 2012; Lenartowicz, Kalar, Congdon, & Poldrack, 2010; K. A. Lindquist & Barrett, 2012; Poldrack & Yarkoni, 2016; Price & Friston, 2005). Some would argue that for
cognitive neuroscience to progress as a science, a serious reconsideration of our existing ontology is in order (Anderson, 2015).

The formal introduction of cognitive ontology as a concept in cognitive neuroscience originated with Price and Friston (Price & Friston, 2005). Motivated by examples in biology, such as the gene ontology consortium (Ashburner et al., 2000), Price and Friston (2005) argued for a comparable project in the cognitive neurosciences. Since Price and Friston (2005), there has been a variety of attempts to either challenge existing ontologies (Anderson, 2015; Klein, 2012; Lenartowicz et al., 2010; Poldrack & Yarkoni, 2016) or derive novel cognitive ontologies from existing neuroimaging or text data (Laird et al., 2015; Poldrack et al., 2012a; Poldrack, Halchenko, & Hanson, 2009; Yeo et al., 2015). Even philosophers of science have taken a keen interest in the cognitive ontology debate in cognitive neuroscience (Anderson, 2015; Hutto, Peeters, & Segundo-Ortin, 2017; Janssen et al., 2017; Klein, 2012). From their perspective, the way we divide up the cognitive domain depends on fundamental assumptions regarding the relationship between cognitive processes and the brain, and the role of taxonomies in the scientific endeavor. Overall, the one thing these views have in common is the dissatisfaction with the current way we “divide up” the cognitive domain. Poldrack (2010) is a recent example:

“Imagine that fMRI had been invented in the 1860s rather than the 1990’s. Instead of being based on modern cognitive psychology, neuroimaging would instead be based on the faculty psychology of Thomas Reid and Dugald Stewart, which provided the mental “faculties” that Gall and the phrenologists attempted to map
onto the brain. Researchers would have presumably jumped from phrenology to fMRI and performed experiments manipulating the engagement of particular mental faculties, or examining individual differences in the strength of the faculties. They almost certainly would have found brain regions that were reliably engaged when a particular faculty was engaged, and potentially would also have found regions whose activity correlated with the strength of each faculty across subjects.”

As noted by Poldrack and Yarkoni (2016), even critics of functional neuroimaging-based cognitive neuroscience have raised concerns about what they view as contrived categories of cognitive processes, such as Uttal (Uttal, 2003):

“Unlike lepidopterists, who have the relatively simple task of gathering and classifying butterflies, psychologists have few such convenient physical anchors. Organizing the myriad proposed psychological components—“butterflies”—of our minds has been and is one of the great unfulfilled challenges of our science. Indeed, it is not only unfulfilled; it has not, in my opinion, been adequately engaged. Rather, hypothetical psychological constructs are invented ad lib and ad hoc without adequate consideration of the fundamental issue of the very plausibility of precise definition. (p. 90)”

Related to debates over cognitive ontology is the emergence of psychological constructionist theories of cognition that are gaining attraction in many areas of cognitive science (Barrett & Satpute, 2013; K. A. Lindquist & Barrett, 2012; Lindquist, Wager,
According to psychological constructionists, the faculty psychology implicit in common-sense distinctions among verbs and nouns in emotional, cognitive and action-based contexts are not respected by the brain. Psychological constructionists argue that this is because mental categories reflected in common-sense vocabulary are not based on distinct mechanisms in the brain and body. Rather, mental events emerge out of more basic psychological operations or “ingredients” (K. A. Lindquist & Barrett, 2012). The goal of psychology and cognitive neuroscience is to discover these psychological ingredients that correspond to operations of the mind at the most basic possible psychological description (below which is simply a biological description; (K. A. Lindquist et al., 2012).

This perceived and empirically supported lack of “naturalness” in the distinctions of our current cognitive ontology have driven some to suggest that current cognitive ontologies should be jettisoned for a more biologically-based one (Anderson, 2015; Klein, 2012). In fact, this impulse has carried over into clinical neuroscience. Many have noted the arbitrariness and the overlap in symptomology of certain psychiatric diagnoses codified in the Diagnostic and Statistical Manual of Mental Disorders (DSM) (Casey et al., 2013). To correct for this perceived lack of objectivity, the National Institute of Mental Health (NIMH) began the RDoC project (Cuthbert & Insel, 2013) to develop a classification of psychiatric diagnoses based on neurobiological dimensions, as well as observable behavior. Thus, this discussion of the status of our current cognitive ontology does not seem to be going away anytime soon. With major funding currently headed in this direction, supported by the NIH (https://datascience.nih.gov/bd2k/about; https://www.nimh.nih.gov/research-priorities/rdoc/index.shtml) and NSF
(https://www.nsf.gov/cise/bigdata/), the debate over ontologies in the cognitive sciences is likely to intensify.

*The Interconnected Role of Task-fMRI and Cognitive Ontology*

The cognitive ontology debate has implications for a task-fMRI ontology, but it’s important to note where these ontological projects diverge. First, a task-fMRI ontology is concerned with categories and divisions that are relevant for activity recorded by task-fMRI, not necessarily cognitive abilities. To assume otherwise would be to assume that the recorded BOLD signal from the brain observed by task-fMRI has the capability of distinguishing the full array of cognitive abilities possessed by human beings. While this is often a working assumption of cognitive neuroscience (the wide variety of behavioral domains studied with fMRI attests to this: http://www.brainmap.org/taxonomy/behaviors.html), this is quite a strong claim. It is certainly possible that task-fMRI activation patterns reflect the activity of much more basic neurological processes shared in common by a wide variety of cognitive abilities, as envisioned by psychological constructionists (K. A. Lindquist & Barrett, 2012).

Second, a task-fMRI ontology is concerned with both latent and concrete categories that are important for describing the task-fMRI experiment. Thus, latent cognitive descriptions of the task-fMRI experiment form one part of a larger description of the task-fMRI environment.

Despite these differences, a task-fMRI ontology and a cognitive ontology inform each other in important ways. For example, identical BOLD activation patterns produced by two classes of tasks claimed to separately involve working memory and inhibition
may cause researchers to jettison the utility of that distinction in a task-fMRI ontology. In terms of a cognitive ontology, this finding may suggest these processes distinguished at the behavioral level might share the same underlying cognitive process. This illustrates an interesting relationship between task-fMRI results and latent features of the task-fMRI experiment: there is a “back-and-forth” between latent features of the task-fMRI experiment and the results of those experiments. For example, researchers in cognitive neuroscience use task-fMRI to localize areas of the brain that correspond to an ostensible cognitive process. On the other hand, failure to localize cognitive processes in task-fMRI experiment may cause researchers to question whether such processes exist as latent constructs. This is not the case for concrete/observable features, such as response type. If we failed to find differences in the pattern of BOLD activity recorded during button presses and joystick movements, we would not then reconsider the utility of their distinction as different actions. Rather, we would conclude that these observable/concrete features do not map onto neural activity recording using task-fMRI, but still may have utility as a distinction in other domains.

This makes latent cognitive processes a tricky feature of a task-fMRI ontology. However, the utility of task-fMRI for cognitive neuroscience is crucially dependent upon its ability to study latent cognitive processes. As described above, many cognitive neuroscientists do not expect the current cognitive ontology to be sufficient for cognitive neuroscience. An ontology of task-fMRI must take this into consideration. Importantly, many arguments over the insufficiency of the current cognitive ontology concern its inability to map to consistent BOLD activation patterns (Lenartowicz et al., 2010; Poldrack & Yarkoni, 2016; Price & Friston, 2005). Thus, the current cognitive ontology
is likely to be insufficient for a task-fMRI ontology. This calls for any ontological project for task-fMRI to provide an objective criterion for what constitutes a “good” latent categorical distinction.

In this dissertation, I apply the same criterion to both latent and concrete potential categories of a task-fMRI ontology: the ability to explain differences in BOLD task-activation patterns. In other words, the utility of categorical distinctions in a task-fMRI ontology is dependent on their ability to explain observed variations in BOLD activation patterns. For example, we have no need for a distinction between latent feature A and latent feature B if their distinction points to no observable difference in BOLD activation patterns. In addition, were concrete features such as stimulus modality, response type, or task paradigm to explain all observable differences in BOLD activation patterns there would be no need to posit latent features to explain these differences. This criterion suggests a data-driven strategy for developing a task-fMRI ontology.

**Strategies for Building a Task-fMRI Ontology**

Even if we agree there is a need for data-driven task-fMRI ontologies, there is no agreed upon way forward. Recent attempts at data-driven task-fMRI ontologies in the cognitive neuroscience literature can be classified into three different categories: 1) latent text-mining approaches (Poldrack et al., 2009; Rubin et al., 2017; Yarkoni et al., 2011; Yeo et al., 2015), classification approaches (Lenartowicz et al., 2010; Poldrack et al., 2009), and activation map similarity matrix approaches (Bolt, Nomi, Yeo, & Uddin, 2017; Laird et al., 2015).
Latent Text Mining Approaches

In recent years, the proliferation of text-mining approaches applied to natural language texts have automated the ontology creation process in many scientific fields. In these approaches, statistical pattern learning algorithms are applied to word counts or other features of a text to derive terms and relationships among them to semi-automate the ontology learning process. These approaches hold promise for a task-fMRI ontology, as they have the potential to identify similar topics across thousands of reported published task-fMRI papers, and estimate the BOLD activation patterns associated with these topics. Recent attempts have been made deriving an ontology of latent cognitive dimensions using a latent text mining approach known as topic modelling (Poldrack et al., 2012b; Rubin et al., 2017; Yeo et al., 2015). At a general level, these studies map latent cognitive dimensions, derived from text documents (Poldrack et al., 2012b; Rubin et al., 2017) or task-paradigm descriptions (Yeo et al., 2015), to patterns of brain activity. These studies provide a data-driven mapping from latent cognitive dimensions to patterns of brain activity. For example, Poldrack et al. (2012) analyzed the occurrence of 605 mental process terms from the Cognitive Atlas Project across 5,809 published task-fMRI articles to estimate latent topics in these texts. They derived 130 separate topics that had been studied across the published articles, related to many common features of the task-fMRI experiment, including movement, emotion, audition, attention and working-memory. Projecting these latent terms onto activation maps reconstructed from peak-activation coordinates reported in the published articles revealed BOLD activation patterns related to each latent term.
While rarely made explicit, this cognitive ontology approach seems to operate on a principle similar to the lexical hypothesis in personality psychology (Mollaret, 2009). Any differences in personality traits or cognitive processes worth making by humans in a linguistic community would be codified into a single- or multi-word phrase. Thus, analysis of self-report questionnaires or natural language texts would reveal the conceptual structure of personality traits or cognitive processing. Regardless of whether this principle is sufficient for a cognitive ontology, it is difficult to see whether this approach would be useful to derive latent features for a task-fMRI ontology. It is certainly possible that latent topics would yield ontological divisions that map on well to differences in BOLD activation patterns. However, latent topic modelling approaches are unlikely to be solely sufficient for a task-fMRI ontology. Latent topic modelling approaches are attempting to model latent terms that explain co-occurrence of words in a text, but not latent terms that sufficiently separate distinct BOLD activation patterns. In effect, the latent text mining approach and task-fMRI ontology approach conceived of above are attempting to maximize different criterion. Thus, latent text mining approaches are unlikely to be solely sufficient for a task-fMRI ontology project.

Classification Approaches

An alternative approach that takes the distinguishability of BOLD activation patterns as its sole criterion is the classification approach to task-fMRI ontology. As in the latent text mining approach, the classification approach is generally used in service of a cognitive ontology, but is just as applicable to concrete features of a task-fMRI ontology. In this approach, activation maps are placed into researcher-hypothesized
cognitive categories (e.g. working memory, inhibition, response selection, etc.) and a classification algorithm (e.g. logistic regression, support vector machine) is trained to distinguish between the activation maps belonging to each category. High classification accuracy of the algorithm is indicative of well-separated categories, and low classification accuracy may be indicative that these categories belong to a single cognitive category. For example, Lenartowicz et al. (Lenartowicz et al., 2010) used a classification approach to categories of “cognitive control”, including “working-memory”, “response inhibition”, “response selection”, and “task-switching”. They found that all categories were discriminable from each other in terms of classification accuracy (greater than 50%), except the “task-switching” vs. “response inhibition” classification and “task-switching” vs. “response selection” selection. In terms of a cognitive ontology, these findings suggest that the latent cognitive category of “task-switching” may not correspond to a unique cognitive process implemented in the brain.

The difficulty with classification as a general approach to constructing a task-fMRI ontology is two-fold: 1) the approach relies on user-specified categories, and 2) it does not explicitly quantify or attempt to explain similarity between BOLD activation maps. To illustrate the first difficulty, imagine scaling up the Lenartowicz et al. (Lenartowicz et al., 2010) classification approach to hundreds of potential latent and concrete categories of a task-fMRI ontology. In this approach, one would progressively remove those potential categories that do not contribute to classification accuracy of the model, until an ideal classification model that balances the number of categories and classification accuracy is obtained. The difficulty of this approach is supplying the category labels of each activation map to the classification algorithm. In terms of
categories like stimulus modality (auditory vs. visual), this would not be too difficult. However, categories such as whether the participant smokes tobacco, whether the participant experiences chronic pain, or participant personality traits are much more difficult to anticipate as important categories for classification. When latent cognitive processes are considered as potential categories, the number of possible categorical distinctions becomes difficult to manage.

The second difficulty is the reliance on classification accuracy as the criterion for a task-fMRI ontology. Classification algorithms, such as logistic regression or linear support vector machines, find a linear combination of the original variables (voxels or brain regions in this case) that maximizes some sort of separability criterion between two categories. However, no information regarding the overall similarity between activation maps of those two categories is provided. This is important for a task-fMRI ontology. For example, we would want two BOLD activation patterns that exhibit the same overall spatial pattern of BOLD activation to be classified into the same category, even if there are small amounts of distinguishing voxel activity between the two patterns.

Analysis of Activation Map Similarity Matrices

Analysis of whole-brain activation map similarity or distance matrices avoids the two identified problems, and I believe better serves the task of constructing a data-driven task-fMRI ontology. Activation-map similarity matrices are constructed by computation of a similarity function between all possible whole-brain BOLD activation maps (Figure 1). In terms of two BOLD activation maps, similarity functions measure the spatial similarity between the two by computation of a similarity function computed across all
voxel or brain region’s activation values. There are a great variety of similarity metrics (often referred to as distance metrics, we use similarity and distance here interchangeably), but commonly used metrics include Euclidean, correlation, cosine, city block, and jaccard metrics. Exploratory analyses of these similarity matrices, such as clustering or latent dimension reduction analyses, provides categorical or dimensional latent features of the task-fMRI environment that explain similarities between whole-brain activation patterns. Rather than attempting to classify between user-specified categories, these analyses estimate latent categories or dimensions that explain the similarity among BOLD activation patterns.

Figure 1. **Activation Map Similarity Matrix.** A hypothetical experiment comparing the degree of activation pattern similarity between the three computed activation maps. This simply involves computing the similarity (e.g. correlation) between all three pairs of activation maps in their voxel-wise activation estimates. From here, we can then begin to build up a similarity matrix for further analysis (warmer colors represent stronger similarity).
Clustering and latent dimension reduction analyses of the activation map similarity matrix provide an effective analysis framework for the task-fMRI ontology envisioned above. The utility of categorical distinctions in a task-fMRI ontology is dependent on their ability to explain *observed* variations in BOLD activation patterns. Clustering solutions and latent dimension reduction analyses of a similarity matrix estimate categorical or dimensional features that *maximize* the explained variance between BOLD activation patterns. In other words, clustering analyses and latent dimension reduction analyses provide an effective framework for deriving a task-fMRI ontology based on the explained variance criterion.

Of course, there are problems with this approach to constructing a task-fMRI ontology as well. For example, the number of clusters or latent dimensions to derive from these algorithms does not have an agreed upon answer. A balance between the amount of explained variance and parsimony is usually suggested, but the definition of balance differs across researchers. In addition, exploratory analysis algorithms provide data-driven clusters or dimensions of similar BOLD activation maps, but do not provide their interpretation. Errors can still be made in the behavioral and cognitive interpretation of these clusters of similar activation maps or latent dimensions. Thus, care is needed in choosing the number of clusters, and their resulting interpretation.

An example of a clustering approach is used by Laird et al. (Laird et al., 2015). In this study, Laird and colleagues were interested in a data-driven estimation of a theoretical cognitive model of face perception. Operating from the same task-fMRI ontology criterion discussed above, Laird et al. (2015) used a hierarchical clustering algorithm applied to an activation map similarity matrix from face discrimination tasks.
They identified four clusters of re-occurring whole-brain activation patterns in face discrimination tasks that mapped onto four interpretable face processing categories. An example of a latent dimension-reduction approach is used by Bolt et al. (Bolt, Nomi, Yeo, et al., 2017). In that study, Bolt and colleagues applied an exploratory factor analysis to a correlation similarity matrix of activation maps and found that BOLD activation patterns are consistently organized around into canonical task-positive/task-negative latent dimension (Figure 2). In addition, this canonical task-positive/task-negative dimension was organized into commonly re-occurring sub-dimensions related to the modality of the stimulus, and the “degree of processing”.

Figure 2. Canonical Task-Positive/Task-Negative Activation Pattern. Heat map of the canonical task-positive/task-negative activation pattern (warmer colors represent activation and cooler colors represent de-activation). This pattern involves an increase in activity in a collection of frontal and parietal brain regions, often including the lateral prefrontal cortex, superior parietal cortex, and posterior medial prefrontal cortex, along with a corresponding decrease in collection of midline frontal and parietal regions (commonly referred to as referred to as the default mode network), including the medial prefrontal cortex and the precuneus/posterior cingulate cortex (Bolt, Nomi, Yeo, et al., 2017).
Analysis Framework for the Dissertation

Within the similarity matrix approach mapped out above, I conducted two empirical studies aimed at assessing a data-driven fMRI ontology. Study one uses the BrainMap database (P. T. Fox & Lancaster, 2002) to assess the quality of existing task-fMRI ontologies, and estimate data-driven categories of a task-fMRI ontology directly from the similarities among activation maps. Study two estimates the contribution of inter-subject variability to both “static” subject-level BOLD activation maps and temporal sequences of BOLD activation maps. Both studies follow the same general set of analysis steps: 1) computation of an activation map * activation map similarity matrix, 2) assessment of existing task-fMRI categories to explain variance between BOLD activation maps using a regression approach, 3) clustering of activation map similarity matrices to derive potential latent categories of a task-fMRI ontology.

I. Computation of activation map * activation map similarity matrices requires the choice of a similarity metric. In study one, a similarity matrix was computed between activation maps derived from a large-scale database of published fMRI experiments, known as the BrainMap database (P. T. Fox & Lancaster, 2002). Given the nature of the activation maps derived from the BrainMap database (discussed in study one), we used a similarity metric known as cosine distance. In study two, we introduce a spatiotemporal similarity metric, known as Dynamic Time Warping (Cuturi, Vert, Birkenes, & Matsui, 2007) to compare
spatiotemporally extended activation patterns in the same manner as BOLD activation maps were compared.

II. The criterion used throughout this dissertation is that the quality of task-fMRI ontology rests in its ability to pick out differences in observed BOLD activation patterns. In other words, the quality of a task-fMRI ontology is proportional to how much variance it explains between similarities in BOLD activation patterns. Up to now, this hasn’t been quantified in any way. To quantify the explained variance of BOLD activation patterns in the second step of our analysis, we use a regression approach extended to distance matrices, known as multivariate distance matrix regression (MDMR; Zapala & Schork, 2006). In MDMR, a distance matrix (Y), representing the pair-wise similarity between all observations is regressed onto a predictor matrix (X) containing predictors for each observation (e.g. gender, IQ, etc.). MDMR is similar to standard regression, except that the predictor matrix (X) is used to partition the sum of squared distances between BOLD activation maps rather than the sums of squares of the voxels or brain regions comprising the BOLD activation maps. The statistics of interest estimated in this approach are the $R^2$ and associated F-statistic. The $R^2$ represents the variance in BOLD activation map similarities explained by the predictor matrix (X). Except in the use of a Euclidean distance function, the F-statistic associated with the $R^2$ does not follow a parametric F-distribution. Thus, a permutation test is needed to assess the significance of each predictor, or set of predictors (discussed below). In study one, we use MDMR to quantify the degree to which
existing task-fMRI categories explain variability among a large set (N > 8000) of group-level BOLD activation maps. In study two, we use MDMR to quantify the degree to which different conditions of a working-memory task explain variability among both “static” subject-level BOLD activation maps and temporal sequences of BOLD activation maps.

III. After the assessment of existing task-fMRI categories using MDMR analysis, we estimate data-driven categories of a task-fMRI ontology directly from the activation map similarity matrix. We do this using a graph clustering approach, known as symmetric non-negative matrix factorization (Kuang, Ding, & Park, 2012). The result of the clustering approach is groupings of similar BOLD activation maps, interpreted to represent potential latent categories of a task-fMRI ontology. To interpret the cluster solution, we use two approaches. First, we calculate the weighted average centroid of each cluster, representing the “representative” BOLD activation pattern of that cluster. Second, we calculate the association between existing task-descriptive categories and the activation maps that belong to each cluster.
Chapter 2: Study One

Background

The exponential growth of task-based functional magnetic resonance imaging research (task-fMRI) has led to issues surrounding the standardization of an ontology or taxonomy to describe and organize task-fMRI results (Hastings et al., 2014; Laird et al., 2015; Poldrack et al., 2011; Turner & Laird, 2012). This lack of standardization poses a problem for the development of representational systems to manage, share and analyze these increasingly large datasets of task-fMRI data. However, good scientific ontologies serve another, more important function: they identify natural divisions or dimensions of the phenomena under study. This function is particularly important for cognitive neuroscience, where cognitive and behavioral features at the psychological level are mapped to the operations of neural function. Thus, an important goal for cognitive neuroscience is the identification of objective categories or dimensions in the task-fMRI environment or person that explain differences in task-fMRI patterns of brain activity.

To address this issue in a quantitative framework, we used a whole-brain representational similarity analysis (RSA) approach (Haxby et al., 2001). Using this approach, we quantified the spatial similarity among whole-brain task-activation maps of the BrainMap database (P. T. Fox & Lancaster, 2002). Using multivariate distance matrix regression (Zapala & Schork, 2006), we predict the variability between task-activation maps by their associated concrete and latent dimensions. The fundamental assumption of this approach is that the dominant dimensions of the task-fMRI experiment are identified in terms of the amount of variance they explain in task-activation map similarities.
Using the spatial similarity estimates among whole-brain task-activation maps we also explore the possibility of estimating latent dimensions directly from these similarities using a graph clustering approach. We take the opposite strategy of recent topic modelling approaches (Poldrack et al., 2012b; Rubin et al., 2017; Yeo et al., 2015). In this approach, latent cognitive dimensions are estimated and then projected onto patterns of brain activity (Poldrack et al., 2012b), or co-occurring latent cognitive dimensions and patterns of brain activity are simultaneously estimated (Rubin et al., 2017; Yeo et al., 2015). In our approach, we reverse this chain of analysis. Commonly re-occurring whole-brain activation patterns, or activation states, are estimated directly from task-fMRI activation maps, and then projected onto cognitive and behavioral descriptors. This approach assumes that the quality of latent cognitive dimensions is determined by how well these dimensions explain differences in task-fMRI activation patterns. Thus, latent cognitive or behavioral dimensions are determined not in terms of their co-occurrence in text, but their co-occurrence with dominant whole-brain activation states. Importantly, these dimensions may or may not map cleanly onto conventional latent cognitive or behavioral descriptors. Rather, these dimensions point the way to potentially new latent cognitive descriptors, that may provide a new technical vocabulary for describing task-fMRI results.

Methods

Outline of Approach

Our approach proceeds in three general steps (Figure 3). First, sparse non-negative matrix factorization (sparse-NMF) is applied to fMRI activation maps reconstructed from experiment-reported peak-activation coordinates to derive a sparse “dictionary” of brain
networks. Spatial similarity estimates between activation maps could be computed at the voxel level, but this ignores spatial dependence between voxels. As opposed to other techniques (e.g., independent component analysis) for estimation of brain networks using the BrainMap database (S. M. Smith et al., 2009; Toro, Fox, & Paus, 2008), NMF provides a data-driven network estimation technique uniquely suited for the data reported in the BrainMap database: all data are positive (i.e. positive activation values) and sparse (i.e. a small number of reported coordinates for each experiment). NMF naturally leads to a parts-based parcellation of the brain given its enforcement of positive weights (Lee & Seung, 1999), even under low signal-to-noise conditions (Lohmann, Volz, & Ullsperger, 2007).

Second, we compute the spatial dissimilarity between the network activation estimates to construct an activation map*activation map dissimilarity matrix. This can be seen as a large-scale whole-brain RSA (Haxby et al., 2001), where dissimilarity estimates are computed between large-scale activity patterns across the entire brain. We then predict the variability among activation maps from the experimenter-labeled task-descriptive categories using multiple distance matrix regression (Zapala & Schork, 2006). MDMR regresses a distance matrix onto a set of continuous or categorical predictors, and has extensive use in behavioral genomics (Schmitz et al., 2011) and studies of individual differences in resting-state functional connectivity (known as connectome-wide association studies; (Shehzad et al., 2014). The parameter of interest from this analysis is the explained variance estimate ($R^2$) for five sets of categorical predictors: analysis decisions, response type, stimulus type, task-paradigm class, and behavioral domain. These estimates provide information regarding what dimensions of the task-fMRI experiment drive the most differences in task-activation maps.
Third, we convert the dissimilarity matrix into a similarity matrix for graph-based clustering. We use a graph clustering algorithm, known as symmetric non-negative matrix factorization (Sym-NMF), that has been shown to outperform commonly-used spectral graph-clustering algorithms (Kuang et al., 2012). Importantly, Sym-NMF allows for overlapping communities of activation maps. As noted above, the goal of this clustering analysis was to estimate potentially overlapping whole-brain activation states. These states may prove useful as candidates for latent cognitive dimensions of a task-fMRI ontology. All code used in study one of the dissertation are provided on the following webpage: https://github.com/NBCLab/neural-driven-cognitive-taxonomy.

*Construction of Activation Maps from BrainMap Coordinates*

At the time of the analysis, the BrainMap database (P. T. Fox & Lancaster, 2002) contained 15900 experimental contrasts from 3216 published manuscripts. Associated with each experimental contrast were the coordinates of statistically significant peak-activation coordinates in MNI152 and Talairach coordinate spaces. Coordinates in Talairach space were converted to MNI152 coordinates using the transform described in Lancaster et al. (2007). The number of contrasts was reduced to ensure that only experimental contrasts of interest were included in the analysis. The only type of experimental contrast included in the analysis was a normal mapping experiment, or exclusively a contrast within healthy participant groups (e.g. no drug treatment). These criteria resulted in a total number of 8919 experimental contrasts for the current analysis.
1) Non-Negative Matrix Factorization for Network Discovery

Voxel \times Activation Map Matrix \rightarrow Network Matrix \rightarrow Network Expression Matrix

2) Prediction of Activation Map Dissimilarity

Activation Map Dissimilarity Matrix \rightarrow Task-fMRI Categories Matrix

3) Graph Clustering for Detection of Overlapping Communities of Activation Maps

Figure 3. Illustration of Data-Driven Approach. 1) First, latent network activation estimates are computed from activation maps derived from the BrainMap database using non-negative matrix factorization (sparse-NMF). NMF decomposes the activation map matrix into a Network Matrix of network voxel weights, and a Network Expression Matrix that reflects the activation estimates of all networks for each map. 2) Next, an activation map dissimilarity matrix is constructed from dissimilarity estimates between latent network
activation patterns for each map. The variability among whole-brain activation patterns is then predicted from experimenter-labeled task-fMRI descriptive categories using MDMR.

3) This dissimilarity matrix is then converted to a similarity matrix and input to a graph clustering algorithm to detect overlapping communities of task-activation maps.

For each of the 8919 experimental contrasts, modeled activation maps were constructed by placing a 12-mm FWHM Gaussian kernel around the center of each peak-activation coordinate reported for each contrast. This kernel size is identical to a previous analysis of the BrainMap database (S. M. Smith et al., 2009). Because the modeled activation maps are extremely sparse (i.e. consisting mainly of zeros), they were sub-sampled by a power of 2 along the X, Y and Z directions for computational feasibility. This sub-sampling resulted in 26459 voxel values per activation map.

**Network Discovery using Non-Negative Matrix Factorization**

To estimate the latent network activation estimates associated with each activation map, we utilized sparse-NMF. The activation maps were first vectorized and placed into a 26459 (voxel) × 8919 (activation map) matrix. This was then input to the sparse-NMF algorithm to model the voxel × activation map matrix as the multiplicative combination of a network matrix and network expression matrix. Each element of both matrices is constrained to be $\geq 0$. The network matrix (voxel × selected number of networks) contains the voxel weights of each network. The network expression matrix (selected number of networks × activation map) represents the latent network activation estimate of each network for each activation map. The NMF algorithm implements the following minimization objective:

$$
min_{W_{\geq 0}, H_{\geq 0}} f(W, H) = \frac{1}{2} \|A - WH\|^2_f
$$

(1)
where A is the voxel × activation map matrix, W is the network matrix, H is the
network expression matrix, and \( \| R \|_F \) is the Frobenius norm of the residual of the W*H
matrix projected onto the A matrix.

The conventional estimation algorithm for NMF estimates elements of the network
and network expression matrix by minimizing the above objective using a multiplicative
update rule (Lee & Seung, 1999), which is computationally infeasible given the
dimensionality of the input matrix (26459 × 8919 matrix). Thus, we used a fast NMF
estimation procedure using the block principal pivoting/active set method described by
Kim and Park (Kim & Park, 2011) and implemented in MATLAB code provided on their
webpage (https://www.cc.gatech.edu/~hpark/nmfsoftware.php). We used a variation of
NMF, sparse-NMF, that adds an additional sparsity constraint imposed on the network
matrix (W matrix) by adding the L1 Norm of the sum of each column of the network matrix
to the minimization objective above (Eq. 1). The estimation algorithm searches for a local
minimum, and is thus initialization-dependent. Rather than using a random initialization of
the network and network expression matrix, we use a singular-value decomposition (SVD)
of the voxel × activation map matrix as the starting initialization of the algorithm described
by Boutsidis and Gallopoulos (Boutsidis & Gallopoulos, 2008), and implemented by
Sotiras et al. (Sotiras, Resnick, & Davatzikos, 2015). The code implemented was provided
on the following webpage: https://github.com/trigeorgis/Deep-Semi-
NMF/blob/master/matlab/NNDSVD.m.

The network expression estimates were used in the subsequent computation of task-
activation map dissimilarity. We chose a higher-order network solution of 70 networks,
comparable to previous analyses of the BrainMap database (Ray et al., 2013; S. M. Smith
et al., 2009). However, the spatial dissimilarity estimates replicated across alternative network solution sizes, as observed by the correlation between dissimilarity estimates of a 70-network solution size with smaller and larger solutions: N = 55, r = 0.91; N = 60, r = 0.93; N = 65, r = 0.96; N = 75, r = 0.95; N = 80, r = 0.94; N = 85, r = 0.92.

**Prediction of Activation Map Dissimilarity with Task-Descriptive Categories**

**Dissimilarity Matrix Computation**

Using the latent network activation estimates of the 70 networks we computed the activation map × activation map dissimilarity matrix with the *cosine* distance metric, representing the dissimilarity in latent network activation estimates between all activation maps. The cosine distance metric was used over other possible distance metrics because it is advantageous for high-dimensional and sparse data (Dhillon & Modha, 2001; Shirkhorshidi, Aghabozorgi, & Wah, 2015). The pair-wise dissimilarity between two activation maps in network expression profiles could vary from 0 (perfect similarity) to 1 (no similarity).

**Task-Descriptive Categories Predictor Matrix Preprocessing**

Categorical predictors fell into five domains: 1) analysis decisions (subtraction vs. baseline contrast, and reported de-activation vs. activation), 2) stimulus type (e.g. visual words, auditory stories, fixation cross, etc.), 3) response type (button press, speech, finger tapping, etc.), 4) paradigm class (encoding, task-switching, counting, etc.), and 5) behavioral domain (action execution, language semantics, working memory). For details of each category, please see: http://www.brainmap.org/taxonomy/. For each domain,
categories were dummy-coded leaving the following variables as the reference variable for each domain: “words” (stimulus type), “write” (response type), “n-back” (paradigm class), and “Perception.Vision.Shape” (behavioral domain).

Because the task-descriptive categories (excluding analysis decisions) allow multi-label classification (i.e. activation maps can have more than one sub-category), multi-label classifications were treated as separate clusters. For example, an activation map classified as Action Inhibition and Attention was treated as belonging to an Action Inhibition/Attention cluster, as opposed to the belonging to both an Action Inhibition and Attention cluster. This increased the number of clusters, and allowed for more precise classifications. A small percentage of activation maps belonged to multi-label categories with fewer than five members (N = 435 for stimulus type; N = 61 for response type; N = 706 for paradigm class; N = 599 for behavioral domain), and these multi-label categories were re-categorized as “unclassified”. Inclusion of multi-label classifications without a membership size cutoff resulted in a heavily overparameterized model (1508 variables), and large sets of perfectly multicollinear predictors (resulting in a rank deficient matrix).

Additional categories were also excluded from the analysis, including those categories with high collinearity with another category (r >0.99; 14 categories), “Not Defined” categories, and multi-label categories that included a “None” classification and a non-None classification. The final dummy coded matrix contained a total of 603 dummy coded variables, including 163 stimulus type variables, 29 response type variables, 215 paradigm class variables, 194 behavioral variables, and 2 analysis decision variables.
Multivariate Distance Matrix Regression

We used multivariate distance matrix regression (MDMR; (Zapala & Schork, 2006)) to model the variability in whole-brain activation map dissimilarity explained by the dummy-coded task-category predictor matrix. Computation of the unique explained variance associated with the five predictors domains was carried out in a leave-one-out fashion. Each domain of categorical predictors was added to the remaining set of categorical predictors to compute the explained variance associated with each set of predictors controlling for all other predictor domains.

Activation maps reported in the BrainMap have a nested structure: multiple activation maps are reported for the same study. To correct for this nesting in the BrainMap database, individual explained variance estimates for each categorical domain were corrected using a permutation test that respected study ID. The permutation test involved a nested reshuffling of activation map labels within each study set. After each reshuffling, an $R^2$ estimate for each set of predictors was computed to construct a null distribution of $R^2$ values. The p-value of the original $R^2$ was then computed as the percentile of the original $R^2$ in the null distribution. The permutation test was run 1000 times.

Graph Clustering Approach

Symmetric non-negative matrix factorization (SymNMF) applies the non-negative matrix factorization framework to pair-wise similarity matrices (Kuang et al., 2012). In the sparse-NMF algorithm described above, the voxel*activation map matrix is factorized into a multiplicative combination of a network matrix (W) and network expression matrix (H). In the symmetric-NMF framework, the symmetric activation map*activation map
similarity matrix \((S)\) is factorized into a cluster membership matrix \((H)\) multiplied by itself. More formally, the symmetric-NMF minimizes the following objective function:

\[
\min_{H \geq 0} \| S - HH^T \|_F^2
\]  

(2)

where \(S\) is the activation map \(\times\) activation map similarity matrix, \(H\) is the cluster membership matrix, and \(\| \cdot \|_F\) is the Frobenius norm of the residual of the \(H^T H\) matrix projected onto the \(S\) matrix. Importantly, SymNMF allows for each activation map to have non-negative weights on more than one cluster. Thus, an activation map can have varying degrees of membership for each cluster. The MATLAB code implement for SymNMF was provided on the following webpage: (https://github.com/andybaoxv/symnmf).

**Determination of the Number of Clusters**

To choose the optimal number of clusters, we used a repeated two-fold cross-validation procedure for cluster sizes ranging from 2 to 20. First, 100 random split-half samples \((n = 4459)\) of the total sample of activation maps were generated. Next, the above SymNMF algorithm was applied to all 100 split-half pairs. Finally, the average pair-wise instability between the 100 split-half cluster centroids was measured using an Amari-type quantity procedure developed by Wu et al. (Wu et al., 2016) (https://github.com/bdgp/staNMF). The average instability, which can vary from 0 (perfect stability) to 1 (no stability), was then plotted across all solutions to search for network solution sizes with low instability. Plotting of the average instability values across all solutions sizes from 2 to 20 revealed two strong local minima in average instability (Figure 3).
**Centroid Activation Patterns and Behavioral Decoding of Clusters**

The resulting community partitions (i.e., clusters) at the two levels of resolution (c = 4 and c = 7) were interpreted in terms of their centroid activation pattern and the BrainMap task-descriptive categories. The centroid activation map of each cluster was computed in two steps: 1) network expression profiles of each activation map (weighted by its cluster membership) were averaged to yield a centroid network expression profile, and 2) the centroid network expression profile was then projected onto voxel space to create the centroid activation map using the following equation:

\[ V = W \times C \quad (3) \]

where \( V \) is 26459 (voxel) \( \times 1 \) centroid activation map, \( W \) is the network matrix (from the NMF solution), and \( C \) is the centroid network expression profile. For visualization, the missing voxel values in the centroid activation map from the sub-sampling procedure described above were interpolated using a penalized least squares procedure (Garcia, 2010) (https://www.mathworks.com/matlabcentral/fileexchange/27994-inpaint-over-missing-data-in-1-d--2-d--3-d--n-d-arrays). The resulting images were then smoothed using a 6mm FWHM Gaussian kernel.

Behavioral decoding of the clusters was performed using the task-descriptive categories used in the MDMR analysis above. In addition to Stimulus Type, Response Type, Paradigm Class, Behavioral Domain and Analysis Decisions, we used the following additional task-descriptive categories: Stimulus Modality, Response Modality, Instruction, and External Variable. Of note, some task-descriptive categories were labeled “None” or “Unknown”, and were excluded from the analysis. To behaviorally decode
each cluster $i$ for each task-descriptive category $s$ of all metadata categories, we computed the forward inference probability normalized by the probability of that cluster:

$$\frac{\text{# of category $s$ instances in cluster $i$}/(\text{# of category $s$ instances across all clusters})}{(\text{# of activation maps in cluster $i$})/(\text{# of activation maps})}$$ (4)

where the numerator represents the $Pr(Cluster \mid Category)$ and the denominator represents the $Pr(Cluster)$.

**Visualization of Behavioral Decoding**

Given the number of task-descriptive categories assessed ($n = 144$; from the above-mentioned categories), we used a data-driven word cloud visualization to illustrate the results. The word cloud visualization displays text with varying sizes and colors throughout space in a visualization window, where word locations are positioned by the “semantic closeness” of each task-descriptive category. The size of the text was proportional to the behavioral decoding result from equation (4). To estimate the “semantic closeness” of the task-descriptive categories we did the following: 1) counted the appearances of all the sub-categories for each activation map to create a 144 (sub-category) × 8919 (activation map) category count matrix, 2) estimated the 144×144 category distance matrix by computing the pair-wise Jaccard distance in the counts between all pairs of sub-categories, and 3) projected the distances between all 144 categories onto a two-dimensional subspace using a nonmetric multidimensional scaling (MDS) algorithm using the mdscale scale function in MATLAB (https://www.mathworks.com/help/stats/mdscale.html). Metadata categories were distinguished by different colors. In many cases, the original word cloud visualization output spatially overlapping category terms. Thus, for clarity, we manually moved spatially overlapping category terms to reduce overlap, while respecting their original position along
the dimensions derived from the MDS solution. We used publically available code for displaying the word cloud visualization, provided on the following webpage:
(https://www.mathworks.com/matlabcentral/fileexchange/53016-wordcloud--classical-).

**Results**

*NMF Results*

In order to estimate the latent network activation estimates underlying each activation map, we applied a sparse-NMF algorithm to the full sample of 8919 task-fMRI activation maps. We derived a high-resolution network parcellation of 70 networks, the same high-model order estimated in previous analyses of the BrainMap database (Ray et al., 2013; Smith et al., 2009). However, the spatial similarity activation estimates replicated across alternative network solution sizes, as observed by the correlation between dissimilarity estimates of a 70-network solution size with smaller and larger solutions: N = 55, r = 0.91; N = 60, r = 0.93; N = 65, r = 0.96; N = 75, r = 0.95; N = 80, r = 0.94; N = 85, r = 0.92. Visual inspection of the 70-network solution (Figure 4) revealed a sparse parcellation of cortical and sub-cortical regions of the brain, corresponding to functionally relevant areas.
Figure 4. **70-Network Parcellation.** Visualization of the 70-network solution derived from the sparse-NMF algorithm. Networks are visualized with separate random colors.

*Prediction of Activation Map Dissimilarity from Task-Descriptive Categories*

Using the latent network activation estimates of the 70 networks we computed the activation map × activation map dissimilarity matrix with the cosine distance metric, representing the dissimilarity in latent network activation estimates between all activation maps. We used MDMR to regress this dissimilarity matrix onto a set of categorical predictors associated with each task-activation map. Categorical predictors fell into five domains: 1) analysis decisions (subtraction vs. baseline contrast, and reported de-activation vs. activation), 2) stimulus type (e.g. visual words, auditory stories, fixation cross, etc.), 3) response type (button press, speech, finger tapping, etc.), 4) paradigm class (encoding, task-switching, counting, etc.), and 5) behavioral domain (action execution, language semantics, working memory).

The BrainMap database contains nested data: multiple task-activation maps are reported for a single study. An unexpected finding of this study was that study ID, the study
from which the task-activation map was reported, accounts for 44.71% of the variance in task-activation map differences. To correct for this nesting in the BrainMap database, individual explained variance estimates for each categorical domain were corrected using a permutation test that respected study ID during permuting of task-activation map labels. Computation of the explained variance associated with the five predictors domains was carried out in a leave-one-out fashion. Each domain of categorical predictors was added to the remaining set of categorical predictors to compute the explained variance associated with each set of predictors controlling for all other predictor domains.

The explained variance accounted for by all predictor domains \( (N_{\text{predictors}} = 603) \) in the model was 19.67%. The explained variance accounted for by analysis decisions (subtraction vs. baseline contrasts, and reported activation vs. de-activations), controlling for other predictor domains, was 0.49% \((p = 0.001)\). The explained variance accounted for by stimulus type, controlling for other predictor domains, was 3.1% \((p = 0.001)\). The explained variance accounted for by response type, controlling for other predictor domains, was 0.57% \((p = 0.014)\). The explained variance accounted for by paradigm class, controlling for other predictor domains, was 4.17% \((p = 0.001)\). The explained variance accounted for by behavioral domain, controlling for other predictor domains, was 3.71% \((p = 0.001)\).

Identification of Re-occurring Whole-Brain Activation States

The full set of available task-descriptive categories explain a limited amount of variance between whole-brain activation patterns. Rather than existing task-descriptive categories, in this section, we explore whether the differences in whole-brain activation
patterns can be explained by a smaller set of latent whole-brain activation states. We first apply a graph clustering approach, symmetric non-negative matrix factorization (Sym-NMF), to derive overlapping clusters of task-activation maps. Overlap allows for the possibility that a single task-activation map may represent a combination of multiple whole-brain activation states. We then project these latent dimensions onto the task-descriptive categories for behavioral and cognitive interpretation.

To choose the optimal number of clusters, we used a cross-validation approach to identify cluster sizes that exhibited stable cluster centroids across split-half samples. Examination of average cluster centroid stability across cross-validation samples revealed two strong local minima of stability at a cluster sizes of four ($M_{stability} = 0.043$) and seven ($M_{stability} = 0.055$) (Figure 5). Thus, further analyses were carried out on these cluster solutions.

![Figure 5. Instability Across Cluster Number.](image_url) Mean instability across 100 random split-half samples for cluster numbers ranging from 2 to 20. Strong global minima were present at a cluster of four and seven (indicated by a vertical red line).
In order to provide a functional interpretation of the clustering results, we computed the centroid activation maps and the prevalence of BrainMap task-descriptive categories (P. T. Fox et al., 2005) in each cluster. The centroid activation maps represent the weighted average activation pattern of the activation maps within the cluster. The association of an experimental descriptor with a cluster was computed as the probability that an experimental descriptor appears given the cluster, normalized by the base probability of the cluster (see Method and Materials).

For the four-cluster solution, the centroid activation maps corresponded to functionally relevant brain areas (Figure 6). The clusters included a visual/dorsal-parietal activation pattern (Four Cluster 1; 4C1), fronto-parietal activation pattern (4C2), a default mode network (DMN) activation pattern (4C3), and an auditory/motor activation pattern (4C4). Consistent with previous reports of domain-general BOLD activation in anterior insula (AI) and dorsal anterior cingulate cortex (dACC) (Bolt, Nomi, Yeo, et al., 2017), activation in the anterior insula (AI) and thalamus was observed across all centroid activation patterns. The task-descriptive categories associated with the four clusters are consistent with previous functional descriptions of these areas of the cortex. 4C1 was associated with paradigms requiring viewing and responding/manipulating visual objects, including “mental rotation”, visual shape perception, and “spatial cognition”. 4C2 was associated with paradigms requiring a variety of complex actions and behaviors, including “working-memory”, “action inhibition” and “language processing”. 4C3 was associated with paradigms requiring recall/recognition of previous information, inferences regarding others’ behavior (“social cognition”), and various affective states (e.g. sadness, fear,
reward, etc.). In addition, consistent with a DMN activation pattern, the activation maps within 4C3 were more likely to be reported as “de-activations”. 4C4 was associated with paradigms requiring overt/covert motor responses, along with listening or responding to simple/complex auditory stimuli.

The seven-cluster solution presented a slightly more fine-grained partition of the four-cluster solution (Figure 7). While 4C2 remained stable from the four-cluster to seven-cluster solution (7C7), the other three clusters from the four-cluster solution were split into two separate clusters in the seven-cluster solution. 4C4 in the four-cluster solution was split into two clusters with activation predominantly in sensory/motor cortices (7C2) and auditory cortices (7C5). Consistent with previous functional descriptions of these areas, the task-descriptive categories associated with 7C2 included “action execution”, flexion/extension and tactile stimulation, and passive listening and pitching monitoring/discrimination for 7C2.

4C3 in the four-cluster solution was split in the seven-cluster solution into two clusters with activation predominantly in the default-mode network (7C4) and sub-cortical structures (basal ganglia and amygdala; 7C6). Consistent with previous functional descriptions of these areas, the task-descriptive categories associated with 7C4 included “theory of mind”, “social cognition”, and “autobiographical recall”, and affective stimuli and “long-term memory” for 7C6.
Cluster 1 ($R^2 = 0.0386$)

Cluster 2 ($R^2 = 0.045$)

Cluster 3 ($R^2 = 0.0329$)

Cluster 4 ($R^2 = 0.0396$)
Figure 6. **Centroid Activation Maps and Behavioral Decoding for Four-Cluster Solution** (BD = Behavioral Domain; PC = Paradigm Class; ExtVariable = External Variable; Stim = Stimulus; Resp = Response). Centroid activation maps with word cloud visualizations (presented to the right each centroid activation map) for the four-cluster solution. Experimental descriptors were sized by the degree of association with each cluster and their positions with respect to each other was determined by their “semantic closeness”. Experimental descriptors were colored according to the metadata category they belong to (e.g. Stimulus Modality, Behavioral Domain, etc.). A color key for each metadata category is provided in the top-right of the figure. The centroid activation maps represent the consistency of activation reported in each voxel across its cluster members (warmer colors represent a greater number of activations recorded at the voxel across activation maps for that cluster). Alongside each cluster label is the amount of unique explained variance that cluster explains in the similarity BOLD activation patterns, controlling for the remaining clusters.

4C1 in the four-cluster solution was split in the seven-cluster solution into visual (7C1) and parietal-frontal clusters (7C3). Activation in 7C1 was predominantly observed in the visual cortex, with slightly weaker activation in the dorsal-parietal and frontal cortices. Activation in 7C3 was predominantly observed dorsal-parietal and frontal cortices, with slightly weaker activation in the visual cortex. Task-descriptive categories associated with 7C1 included passive viewing paradigms, including “action observation”, and visual shape perception, and perception of moving objects. Task-descriptive categories associated with 7C3 included active viewing paradigms, including “mental rotation”, “spatial cognition”, and delayed match to sample tasks.

An important observation apparent in the clustering solution is that task-descriptive categories “distant” in semantic space (farther apart in the word cloud) load strongly onto single activation states. In other words, descriptive categories that are semantically “distant”, are often not so “distant” in the latent BOLD activation space. For example, consider the following two task paradigms: the Stroop color word task, and an overt word generation task. The psychological processes generally inferred to underlie these tasks
include action inhibition (Stroop, 1935) for the Stroop color word task, and lexical processing (Friedman et al., 1998) for a word generation task. However, the whole-brain activation patterns of these paradigms are similar, observed by the fact they both load strongly onto the same latent whole-brain activation state: 4C4 in the four-cluster solution and 7C7 in the seven-cluster solution. This suggests the neural response measured by task-fMRI is similar for the Stroop color word task and word generation task. Thus, whole-brain activation states may be informative in terms of mapping the similarity in functional anatomy between different cognitive or behavioral processes.
Figure 7. *Centroid Activation Maps and Behavioral Decoding for Seven-Cluster Solution* (BD = Behavioral Domain; PC = Paradigm Class; ExtVariable = External Variable; Stim = Stimulus; Resp = Response). Centroid activation maps with word cloud visualizations (presented to the right each centroid activation map) for the four-cluster solution.

*Whole-Brain Activation State and Task-Descriptive Category Comparison*

To compare the explained variance from these latent whole-brain activation states to the task-descriptive categories, we conducted an MDMR analysis regressing the weighted membership coefficients of the four- and seven-cluster separately onto the activation map * activation map dissimilarity matrix. The total explained variance for the four-cluster solution and seven-cluster solution were 20.95% ($p = 0.001$) and 34.37% ($p = 0.001$), respectively. Thus, both lower-order latent activation states ($N = 4$ and $N = 7$) explained a greater amount of variance in task-activation map dissimilarity than the higher-order descriptive categories ($N = 603$). This difference in explained variance can be visualized with a re-ordered activation map*activation map similarity matrix according to either the latent activation states or two example task-descriptive categories (*Behavioral Domain* and *Paradigm Class*; Figure 8).
Figure 8. **Comparison of Latent Activation State Clusters and Task-Descriptive Category Clusters.** To visualize the differences in explained variance between the latent activation states and two task-descriptive categories (*Behavioral Domain* and *Paradigm Class*), we visualized a re-ordered activation map * activation map similarity matrix according to the latent activation state and task-descriptive categories. The re-ordered similarity matrices for the latent activation state and task-descriptive categories were created by re-organizing each activation map in the similarity matrix according to their dominant whole-brain activation state cluster (i.e. assignment of each map to the cluster with the strongest loading) or *Behavioral Domain/Paradigm Class*. The values in the similarity matrices vary from 0 (no similarity; cool colors) to 1 (max similarity; warm colors). The red lines organized along the diagonal of the similarity matrix trace groups of activation maps that belong to a single cluster according to their latent activation state or *Behavioral Domain/Paradigm Class* category. Similarity matrices with brighter colors within the red cluster lines, and cooler colors outside, correspond to more compact and well-separated groups of task-activation maps.
Discussion

Much cognitive neuroscience research in the past two decades has attempted to map latent cognitive processes to neuroanatomy. However, another driving goal of cognitive neuroscience is to generate or constrain theories of cognitive processes from functional neuroimaging data. Directed at this latter goal, the primary aim of this study was to identify the dominant dimensions upon which whole-brain BOLD activation maps vary. In other words, what dimensions of the task environment is the BOLD response most attuned to? Task-based fMRI research has conventionally used a fluid set of “task-descriptors” to interpret the differences between activation maps, including both concrete (stimulus modality, response modality and task paradigm) and latent (behavioral domain) descriptors. Using a novel analysis pipeline applied to the BrainMap database, we quantify the variance explained by these concrete and latent descriptors. The full model of task-descriptive categories (N = 603 predictors) explained a moderate proportion of variance ($R^2 = 0.197$) in the similarities between activation maps. We find that stimulus type, paradigm class and behavioral domain uniquely explain a small ($R^2 < 0.05$), but non-trivial amount of variance in similarity between activation maps.

It is difficult to assess these explained variance estimates in terms of absolute value, as the disparate quality of the activation maps (derived from peak-activation coordinates) place an unknown upper limit on the variance explainable by the task-descriptive categories. However, we can make relative comparisons. We find that the concrete domain of Paradigm Class (or task paradigm), explains the greatest amount of variance among the task-descriptive categories, controlling for other domains, ($R^2 = 0.417$). This suggests that differences in task experiments drive the most observed
differences in fMRI task-activation patterns over and above differences other domains (e.g. behavioral domain). Thus, task-fMRI researchers should be aware that the choice of experimental task may be of more crucial importance than whether they require the same supposed cognitive or behavioral process. An unexpected finding of this study is that a substantial portion of variance in the similarity between activation maps can be explained by the study ID ($R^2 = 0.447$); over two times greater than the full set full model of task-descriptive categories ($R^2 = 0.197$). This can be for multiple reasons: 1) differences in peak-coordinate reporting, 2) differences in pre-processing pipelines and scanner hardware, or 3) differences in significance thresholding procedures (Eklund, Nichols, & Knutsson, 2016). These differences in study ID may be reduced with the sharing of unthreshold statistical maps (Bolt, Nomi, Yeo, et al., 2017; Gorgolewski et al., 2015), that avoids peak-coordinate reporting and thresholding procedures.

We assume that the quality of latent cognitive and behavioral descriptors for a task-fMRI ontology is contingent upon how well these descriptors explain differences in BOLD activation patterns. Thus, we attempt to derive a set of latent activation states and associated cognitive/behavioral descriptors that maximize the explained variance between BOLD activation patterns. With just a set of four (or seven) latent activation states that map on well to sensible combinations of behavioral domains, we can explain a greater amount of variance in BOLD activation maps than the full set of task-descriptive categories ($N = 603$). One might argue that this is trivial: one would expect that a solution built to maximize explained variance between activation patterns will of course explain more variance than the *a priori* task-descriptive category predictors. However, we do not believe this is trivial for two reasons: 1) though it is trivial fact that all clustering
algorithms at a general level estimate clusters that maximize explained variance, it is not
trivial that a \textit{low-dimensional} clustering solution (N= 4 or N = 7) can explain greater
variance than the high-dimensional task-descriptive categories (N = 603). Relatedly, 2)
explained variance of activation map dissimilarities \textit{should} be our criterion for what
constitutes an adequate task-fMRI ontology. For example, if a four-cluster solution
explained 99\% of the variance in task-activation map dissimilarity, adding more task-
descriptive categories amounts to an unnecessary overparameterization of this space.

The task-descriptive categories associated with the latent activation states from
the 4-cluster and 7-cluster were consistent with their associated activation pattern. The
four latent activation states can be generally described as object-viewing (4C1),
inhibition/control/language (4C2), self-memory/affective/social (4C3), and auditory-
stimulus/motor-action (4C4) states. The seven latent activation states can be generally
described as passive object-viewing (7C1), sensory-motor (7C2), active object-viewing
(7C3), social/theory of mind (7C4), auditory (7C5), affective (7C6) and
inhibition/control/language (7C7). Examination of the four-cluster and seven-cluster
solutions centroid activation patterns revealed a set of regions commonly activated across
the BrainMap database. Consistent activation was observed across regions that make up
the traditionally described “task-positive” and “task-negative” activation/de-activation
pattern (Bolt, Nomi, Yeo, et al., 2017; M. D. Fox et al., 2005; Raichle et al., 2001). The
“task-positive” activation pattern traditionally includes elements of the fronto-parietal
(DLPFC, SPC and DMPFC) and salience networks (dorsal anterior cingulate cortex and
AI). Consistency of activation was particularly predominant in the AI and dorsal anterior
cingulate cortex (dACC) across all centroid activation maps, in accord with previous
studies demonstrating the AI and dACC act as brain network hubs (Bolt, Nomi, Rubinov, & Uddin, 2017; Cole et al., 2013; Uddin, 2015). The “task-negative” de-activation pattern, generally restricted to the DMN (Raichle et al., 2001), was represented in both the four- and seven-cluster solution. Consistent with the “task-negative” ascription, the activation maps of these clusters were more likely to be reported as “de-activations”.

Related to a task-fMRI ontology, is the more fundamental idea of a cognitive ontology. Recent research has suggested that that the currently available latent “cognitive” descriptors do not reflect natural divisions of neural function (Anderson, 2015; Cuthbert & Insel, 2013; Poldrack & Yarkoni, 2016). However, we avoid equating the discovered whole-brain activation states with potential categories of a cognitive ontology for two reasons. First, mapping this clustering solution to taxonomic categories of cognition assumes that the BOLD contrast used in fMRI has the capability of distinguishing the full array of cognitive abilities possessed by human beings. Second, a purely “neuro-centric” cognitive ontology would miss out on the bodily and environmental factors also constitutive of proper cognitive functioning (Anderson, 2003; Chemero, 2011; Hutchins, 2010). Thus, while these latent activation states may be important for understanding the dominant cognitive and behavioral dimensions explaining differences in BOLD activation patterns, they by no means exhaust the potential categories of a cognitive ontology.

Limitations

One objection to the current approach is that it is circular. According to this objection, the use of conventional task-descriptive categories of cognitive functioning
and task paradigms for interpretation in the behavioral decoding analysis essentially reifies these categories in our neural-driven categorization. Thus, the analysis relies on a circular inference. In response, we agree that the mix of ordinary and technical terms codified in the BrainMap experimental descriptors is relied upon in our descriptions of the data-driven categories, but reliance on this terminology does not constitute circularity. The behavioral decoding analysis maps clusters to their most representative experimental descriptors, but does not identify the clusters with those experimental descriptors. For example, descriptions of complex physics topics often rely on ordinary linguistic terminology, but this description does not then identify the physical systems themselves with these ordinary terms. In the same manner, describing the data-derived categories in terms of their association with ordinary or current technical terms of cognitive science does not constitute circularity. By examining the pattern of experimental descriptors associated with each cluster, we may find pointers to an underlying neural process that unifies these descriptors.

Due to the disparate representation of activation maps in the BrainMap database in terms of peak-activation coordinates, the precision of the distinctions between possible clusters is limited. However, the substantial sample size of reported task-fMRI experiments provided by the BrainMap database is unmatched in terms of power and detailed experimental metadata. The increasing size of databases containing unthresholded activation maps, such as NeuroVault (Gorgolewski et al., 2015), offers the potential for a more precise or fine-grained categorization using the approach applied in this study. In addition, the task-activation maps in the BrainMap database were computed using a variety of analytic approaches and experimental designs, which introduces an extra source of
variability among the activation maps that cannot be accounted for in our analysis. Nevertheless, we hope the results presented here provide a starting point for a well-developed task-fMRI ontology.
Chapter 3: Study Two

Background

Study two aims to extend the results of study one in two ways. First, group-level activation maps were the units of analysis in study one. Thus, no information was available regarding inter-subject variability between BOLD activation patterns. In study two, we use subject-level BOLD activation maps to estimate the contribution of inter-subject variability. Second, we extend the analysis of similarity between whole-brain activation patterns to spatiotemporal activation patterns, using a distance metric known as Dynamic Time Warping (DTW).

In study one it was discovered that a substantial proportion of variance in task-activation map dissimilarity could be explained by the study from which the map was obtained. Much of this study variability is likely to represent uninteresting differences in study-specific factors, such as scanner hardware, preprocessing pipelines, or peak-coordinate reporting. However, this does point to the fact the majority of variance we find in BOLD activation maps is not likely to be due to experimental manipulations of the task fMRI design. One source of variance known to be ubiquitous across both task fMRI (Miller et al., 2002, 2009) and resting-state fMRI (Finn et al., 2015; Rosenberg, Finn, Scheinost, Constable, & Chun, 2017) is inter-subject variability. For example, Miller et al. (2009) found that whole-brain activation patterns from two different tasks of a single subject were more similar than whole-brain activation patterns from different individuals performing the same task. These findings suggest that inter-subject variability is likely to be an important driver of differences in BOLD activation maps. Thus, in study two we
examine the contribution of inter-subject variability to differences in whole-brain BOLD activation patterns.

In addition to “static” activation patterns, study two aims to also examine inter-subject variability of spatiotemporal activation patterns. It is important to remind ourselves of what the BOLD “activation” map represents. These “activation” maps are simply beta coefficient estimates representing the degree of association between the BOLD time series of each voxel and the time course of the task blocks or events of interest (see Appendix A). Depending on how we model the time course of the task blocks or events of interest (e.g., convolve the task design with a canonical hemodynamic response function or use a more model free approach, such as the finite-impulse-response model; (M. A. Lindquist, Loh, Atlas, & Wager, 2009), we are going to lose some of the information regarding the dynamics of activation in response to the task block or event. This spatiotemporal information is likely to be important. In the same manner that the operation of other organs of the body exhibits commonly re-occurring spatiotemporal patterns (e.g. the heart), it is likely that the brain operates in a similar manner. To capture these dynamics of whole-brain activation patterns, we use a time series similarity metric well-known in speech processing, face recognition and time-series classification and clustering known as dynamic time warping (Cuturi et al., 2007; Izakian, Pedrycz, & Jamal, 2015; Petitjean, Ketterlin, & Gançarski, 2011; Rath & Manmatha, 2003).

These analyses are conducted on subject-level BOLD activation patterns from a working-memory task that systematically varies stimulus and cognitive load across conditions. In the first analysis, we assess the variability of “static” subject-level BOLD activation maps across both experimental manipulations and subjects. This is a subject-
level implementation of the approach pursued in study one. In the second analysis, we introduce a spatiotemporal similarity metric for subject-level BOLD activation patterns on the same task, and illustrate the results with this novel similarity metric. In a similar approach to study one, we also derive data-driven spatiotemporal activation patterns using a clustering approach applied to the spatiotemporal activation pattern similarities.

Methods

fMRI Data

Neuroimaging data from 100 unrelated, healthy, right-handed adults (Age Ranges: 22-25: 46, 26-30: 54, 31-35: 48, 36+: 2; 75 female) made available through the Human Connectome Project (HCP) 2014 release were used for this study. Participants were recruited from the surrounding area of Washington University (St. Louis, MO). All participants gave informed consent before participating in the study, as described in Van Essen et al. (Van Essen et al., 2013).

Task Description

The working-memory task was a N-back task that involved visual processing of category specific visual stimuli. The entire task consisted of two runs, that each included $\frac{1}{2} 0$-back (a target visual cue is presented at the beginning of each block and the participant is directed to respond when that visual stimulus is presented) and $\frac{1}{2} 2$-back blocks (the participant is directed to respond when the current visual stimulus is identical to one presented two presentations back) that presented one of four stimulus types blocks (body parts, faces, places and tools). Thus, there were eight 25s task blocks for each run.
representing all possible combinations of task type (0-back and 2-back) and stimulus type (body parts, faces, places and tools), along with four 15s visual fixation blocks. At the start of each block, there is a 2.5 sec visual cue indicating the task type (0-back versus 2-back), followed by ten 2.5s trials, where a stimulus was presented for 2s, followed by a 500ms inter-trial interval.

*Static Activation Maps*

For analysis of subject-level “static” activation maps, we used *unthresholded* subject-level BOLD activation maps for each of the eight possible task blocks (2-back/face, 0-back/face, 2-back/tool, etc.). These maps were provided directly from the HCP, and no preprocessing or analysis was necessary to compute the subject-level BOLD activation maps. The subject-level BOLD activation maps were *z*-scored beta images representing subject-level BOLD activity in response to the eight possible task blocks (see Appendix A for details of how activation maps are created).

*Multivariate Distance Matrix Regression of Subject-Level Activation Maps*

In order to assess the variance in subject-level BOLD activation maps due to stimulus conditions and subjects we used multivariate distance matrix regression (MDMR; (Zapala & Schork, 2006)). We computed the correlation distance between each subject-level BOLD activation map to create an activation map dissimilarity matrix. We then regressed this distance matrix onto three predictors, subject ID (treated as a fixed effect), stimulus condition (body, face, place and tool), and 0-back vs. 2-back condition.
To correct the stimulus and 0-back/2-back for nesting of activation maps within each subject, a nested permutation test was used respecting subject ID (similar to the approach in study one). The permutation test involved a nested reshuffling of activation map labels within each subject. After each reshuffling, an $R^2$ estimate for each was computed to construct a null distribution of $R^2$ values. The $p$-value of the original $R^2$ was then computed as the percentile of the original $R^2$ in the null distribution. The permutation test was run 1000 times.

*Dynamic Time Warping Similarity Measure*

To extend the task-fMRI ontology approach to spatiotemporal BOLD activation patterns, we apply a measure known as dynamic time warping (DTW; Figure 9) to the multivariate time series within each of the 16 task blocks (each of the 8 task blocks are presented twice across both task runs) for each subject. As describe above, the multivariate time series of each task block consisting of both a 2.5s visual cue and 25s task portion (preprocessing of the original scans is described below).

Conventional dynamic time warping (DTW) algorithms fail to satisfy the triangular inequality, which is necessary for most clustering applications. As illustrated in Figure 9, conventional DTW algorithms select the optimal path between two sequences and its associated score as a measure of similarity. Thus, we used a measure known as the dynamic global alignment kernel (Cuturi et al., 2007). DGAK computes the smoothed version of the maximum of the score spanned by all possible paths, resulting in a positive definite kernel.
Figure 9. **Illustration of Dynamic Time Warping Similarity Measure.** A) Imagine a hypothetical example where we are comparing the multivariate time series between two task blocks with a length of 5 seconds, and a sample (TR) every second. Displayed is the whole-brain BOLD estimate of each voxel at each second in colors (warmer colors represent higher BOLD activity levels). Notice that the multivariate time series are very similar: the first three time points exhibit very similar activation patterns with divergent patterns thereafter: the block one-time series exhibits a visual cortex activation pattern at time 4 and time 5, while the block two-time series stays constant from time 3 to time 4, and then exhibits a visual cortex activation at time 5. So, overall, the blocks seem to be very similar in their spatiotemporal activity except at time 4. The first step in the dynamic time warping similarity measure is to build up a dissimilarity matrix of activation pattern similarity between the whole-brain BOLD estimates between each time point of each block (represented by the double-headed blue arrows) and the resulting dissimilarity matrix is illustrated to the right. B) The objective of the dynamic time warping algorithm is to find the optimal path through the matrix starting at time 1 (top left) and ending at time 5 (bottom right) that minimizes the sum total distance along the path (satisfying certain constraints – can’t move backwards, etc.) Moving through the matrix from time 1 to time 3 we move right down the diagonal as the activation patterns from block one and two are very similar at those time points. We then move from horizontally through the matrix at time 3 in block one to time 4 in block two reflecting the fact that the activation pattern at time 3 in block one is more similar (i.e. least dissimilar) to the activation pattern at time 4 of block two. We then move vertically down from there reflecting the fact that the activation pattern at time 4 in block one is more similar to the activation pattern at time 5 in block two. And finally, we move to the final time point. The sum of the distances along this path reflects the “distance” between the two blocks in their multivariate time series, and forms the basis for constructing the dissimilarity/similarity matrix between all pairs of blocks.
Data Preprocessing for Dynamic Time Warping Analysis

For the dynamic time warping analysis, we used minimally preprocessed data provided through the HCP. The minimal preprocessing pipeline involved gradient distortion correction, motion correction, registration to the Montreal Neurological Institute (MNI) template, and intensity normalization. The details of the minimal preprocessing pipeline are described in (Glasser et al., 2013). Additional preprocessing steps included demeaning, variance normalization (normalizing the data to their standard deviations from the mean, z-score) and concatenation of the time series from the right-left and left-right runs, respectively for each task using the Connectome Workbench (Marcus et al., 2011). Data was resampled to 3mm voxel size for computational feasibility, time courses were despiiked using AFNI’s 3dDespike, an interpolative scrubbing procedure, and time-series were then detrended using FSL’s nonlinear Gaussian weighted least-squares fitting with a cutoff of 100s (S. M. Smith et al., 2004). The functional data were then spatially smoothed (5mm full width at half maximum) with FSL, and nuisance covariate regression was performed (Friston’s 24 motion parameters, namely each of the 6 motion parameters of the current and preceding volume, plus each of these values squared, ventricle and white matter signals) using the Data Processing and Analysis for Brain Imaging (DPABI) toolbox (Yan, Wang, Zuo, & Zang, 2016). No participants included in the analysis displayed gross motion (relative Root-Mean Squared-Framewise Displacement (Jenkinson, Bannister, Brady, & Smith, 2002); RMS-FD < 0.55mm; (Satterthwaite et al., 2013).
Multivariate Distance Matrix Regression

To account for spatial dependencies among voxels, we extracted time series from the high-resolution Schaefer et al. (Schaefer et al., 2017) parcellation, which was found to be more homogeneous than other available ROI parcellations. Voxel time scores within the regions of interest (ROIs) of the parcellation were averaged together. DGAK was computed between the multivariate ROI time series from each of the 16 task blocks for all 100 participants to create a spatiotemporal activation map dissimilarity matrix. As in the first analysis, we then regressed this distance matrix onto three predictors, subject ID (treated as a fixed effect), stimulus condition (body, face, place and tool), and 0-back vs. 2-back condition. The same nested permutation framework in the first analysis is also applied to derive corrected $p$-values.

Graph Clustering of Spatiotemporal Activation Similarity Matrices

To derive latent spatiotemporal activation patterns from the spatiotemporal activation maps, we applied Symmetric-NMF (SymNMF) to the similarity matrix of spatiotemporal activation patterns. To choose the number of clusters, we plotted the explained variance by the number of clusters (Figure 11). Based on this criterion we chose a cluster solution with two clusters. To visualize the latent spatiotemporal activation pattern associated with each cluster, we computed the centroid spatiotemporal pattern by averaging together all multivariate time series ($N = 800$ ROIs) from all 16 task blocks, weighted by its cluster membership weight. For visualization, we project the
multivariate time series of each ROI back onto voxel space, and spatially smooth the voxel values across ROIs with a Gaussian kernel (FWHM = 6mm).

**Behavioral Decoding of Clusters**

In a similar manner to study one, we behaviorally decode each spatiotemporal activation cluster by counting the proportion of categories (0-back, 2-back, body, face, place, and tool) belonging to the spatiotemporal activation patterns within each cluster. Of note, because the number of spatiotemporal activation patterns in each cluster was close to equal (N = 795 for cluster one, and N = 805 for cluster two), we did not correct the proportions for the baseline probability of the cluster as we did in study one.

**Results**

**Prediction of Activation Map Dissimilarity**

For the analysis of inter-subject variability of BOLD activation patterns, we used subject-level BOLD activation maps from each of the 8 task block types of the working-memory task, including 0-back body, 0-back faces, 0-back places, 0-back tools, 2-back body, 2-back faces, 2-back places, and 2-back tools. We computed the activation map × activation map dissimilarity matrix between all 8 task block types for all subjects, representing the dissimilarity between all subject-level BOLD activation maps. We then used MDMR to regress this dissimilarity matrix onto subject (modelled as a fixed effect), stimulus type, and 0-back vs. 2-back condition.

The explained variance accounted for by the full model was 81.56%. The dominant proportion of variance was accounted for by subject, at 75.88%. The
dominance of the subject effect can be visualized as a strong block pattern along the diagonal of the activation map * activation map similarity matrix (Figure 10). The explained variance accounted for by the 2-back vs. 0-back condition was 2.69% ($p = 0.001$). The explained variance accounted for by stimulus type was 5.68% ($p = 0.001$).

**Prediction of Spatiotemporal Activation Pattern Dissimilarity**

For the analysis of spatiotemporal activation patterns, we used the 16 task blocks (2 runs of each the 8 task block types) of the working-memory task. DGAK was computed between the multivariate time series within each of the 16 task blocks for all subjects to create a spatiotemporal activation pattern × pattern dissimilarity matrix. We then used MDMR to regress this dissimilarity matrix onto subject (modelled as a fixed effect), stimulus type, and 0-back vs. 2-back condition.

The explained variance accounted for by the full model was 17.61%, indicating that spatiotemporal activation patterns are more heterogeneous within and across subjects than static activation patterns. Similar to static BOLD activation maps, the dominant proportion of variance was accounted for by subject, at 10.04%. Interestingly, the explained variance accounted for by the 2-back vs. 0-back condition was 3.55% ($p = 0.001$), greater than that in the static BOLD activation maps above. The explained variance accounted for by stimulus was 3.9% ($p = 0.001$).
Figure 10. **Static Activation Map Dissimilarity Matrix.** Similarity matrix between all subject-level BOLD activation maps (top), and a zoomed in view of the same similarity matrix (bottom). Values of the correlation similarity metric (0 – 2) are mapped to color, with brighter colors representing higher similarity and cooler colors representing lower similarity. To illustrate the dominance of inter-subject variability in subject-level BOLD activation maps, we present a zoomed in view at the bottom of the figure (outlined in red in the original similarity matrix). Eight activation maps are used for each subject, and activation maps belonging to one subject are adjacent to each other in the similarity matrix. As is visually apparent, much of the variance between activation patterns is explained by consistent inter-subject differences.
Clustering of Spatiotemporal Activation Patterns

To extract latent spatiotemporal clusters from the spatiotemporal activation pattern similarities, we convert the activation pattern dissimilarity matrix to a similarity matrix and apply symNMF. To choose the number of clusters, we plotted an explained variance by cluster number plot (Figure 11). Visual examination of the plot revealed that the one- and two-cluster solutions significantly add to the proportion of variance explained, but beyond a two-cluster size, the additional explained variance levels off and begins to decrease. Thus, we chose a cluster solution of two. The explained variance in spatiotemporal activation dissimilarity accounted for by the two-cluster solution was 83.94%. Reorganizing the similarity matrix according to the cluster structure makes this explained variance in similarity patterns visually apparent (Figure 12).

![Explained Variance by Cluster Number](image)

**Figure 11. Explained Variance by Cluster Number.** Explained variance associated with each possible cluster size (2 – 20). As can be observed from the plot, cluster numbers one and two explain a substantial portion of the variance (83.94%), after which the explained variance for each additional cluster size drops off and begins to exponentially decrease after a cluster number of 3.
Figure 12. **Spatiotemporal Activation Pattern Similarity Matrices.** Two similarity matrices are presented: 1) Left: similarity matrix between spatiotemporal patterns not organized by cluster structure, and 2) Right: the same similarity matrix organized by the two-cluster solution. Interpretation of the values in the similarity matrix are the same as in Figure 10.

Visualization of the spatiotemporal cluster centroids revealed two distinct spatiotemporal activation patterns in response to the task (Figure 13). Two general differences are observed in the temporal sequence of activation patterns for cluster one and cluster two. First, the spatiotemporal activation pattern for cluster one starts with default-mode network (DMN) de-activation and moves to a task-positive network activation. The spatiotemporal activation pattern for cluster two starts with a DMN activation and moves towards a task-positive activation pattern at a much slower pace than that observed in cluster one. Second, while the cluster one spatiotemporal pattern moves into task-positive de-activation in the latter half of the task block, the spatiotemporal pattern maintains its task-positive activation throughout the block.
This difference between the centroid spatiotemporal patterns of cluster one and two may be explained in terms of the condition categories (stimulus or 0-back/2-back condition) they predominantly occur in. Thus, we computed the proportion of each condition category that occurs in either cluster one or cluster two. Stimulus categories (body, face, place and tools) occurred in almost equal proportions for cluster one (body=0.55, face=0.48, place=0.57, tool=0.42) versus cluster two (body=0.45, face=0.52, place=0.43, tool = 0.58). However, the 0-back and 2-back condition categories occurred in unequal proportions for cluster one (0-back=0.74, 2-back=0.25) versus cluster two (0-back=0.26, 2-back=0.75). This is consistent with the observed temporal sequence of activation patterns for the cluster centroids. Specifically, the 2-back and 0-back conditions require quite a different degree of “cognitive load” or effort from the participant. The 0-back condition, most associated with cluster one, requires a very small short-term memory demand, while the 2-back, most associated with cluster two, requires a much greater short-term memory demand. This may explain why the cluster two centroid spatiotemporal activation pattern (most associated with 2-back) maintains a task-positive activation pattern, as the task-positive activation pattern is often associated with task-directed attention (Bolt, Nomi, Yeo, et al., 2017).
Figure 13. **Spatiotemporal Centroids for Two-Cluster Solution.** Temporal sequence of activation patterns associated with cluster one (to the left) and cluster two (to the left). BOLD values for each voxel are presented in blue for negative values (de-activation), and red for positive values (activation). Rather than visually presenting all activation patterns in the temporal sequence of each centroid (39TRs/samples for 25s), we present the activation maps corresponding to intervals of 5 TRs.

**Discussion**

Overall, the goal of study two was to assess the contribution of inter-subject variability within a task-fMRI ontology. The specific aims of study two were 1) to assess the degree of inter-subject variability in “static” subject-level BOLD activation maps, and 2) spatiotemporal activation patterns. To address these aims, we used subject-level
BOLD activation maps from a visual working-memory task. Regarding aim one, the results suggest that inter-subject variability is a large driver of differences in BOLD activation patterns. In fact, the variability due to subjects “swamps” the variability due to task-demand and stimulus changes in the task environment: 75.88% to 2.69% and 5.68%, respectively. Regarding aim two, we introduce a similarity metric for comparing spatiotemporal patterns (DTW), and observe that spatiotemporal patterns are more heterogeneous than “static” activation maps both within and across subjects. We also discovered two dominant spatiotemporal activity patterns across the 8 task block types that corresponded roughly to the division between 0-back and 2-back working-memory conditions.

Inter-subject variability in fMRI functional connectivity patterns is currently a predominant research topic in cognitive neuroscience (Finn et al., 2015; Rosenberg et al., 2017). For example, Finn et al. (2015) has demonstrated that functional connectivity patterns of an individual subject act as a “fingerprint” that can be used to identify that subject regardless of what state he/she is in (e.g. performing a task or resting). Inter-subject variability has also been studied in BOLD activation patterns recorded during task performance with similar conclusions (Miller et al., 2009). The results of study two are consistent with these previous studies. However, we are the first, to our knowledge, to provide a quantitative estimate of the degree of inter-subject variability in subject-level BOLD activation patterns. The explained variance estimate is very large: 75.88% of the observed variance in BOLD activation map similarities. Another question is whether this observed variation corresponds to functional variation in neural processes. Given the BOLD contrast is inherently related to the vasculature and its pattern of innervation, it is
entirely possible that inter-subject variability in BOLD activation patterns is not due to differences in neural processing, but differences in vascular innervation of the neural tissue. It is likely that both contribute to inter-subject variation in BOLD activation maps, but future research is needed to disambiguate these possibilities.

The BOLD activity we record from the brain during a task includes both spatial and temporal information. Study one and the first analysis of study two collapse across time and analyze spatial patterns. However, this approach is bound to miss important spatiotemporal patterns in response to the task. To extend whole-brain similarity analyses beyond “static” activation patterns to spatiotemporal activation patterns, we use a time-series similarity metric known as DTW. Applying the same MDMR analysis to a similarity matrix created using DTW, we see a similar dominance of inter-subject variability in our results ($R^2 = 0.10$). However, the overall variance explained by subjects and task conditions was significantly lower than the static activation maps. This might be due to heterogeneity in the spatiotemporal activity that is related to functional differences in neural processing. However, it’s also possible that this heterogeneity is due to the increased number of variables being compared (sequences of activation maps, as opposed to a single activation map), which may result in more noisy similarity estimates. In this case, within-subject averaging of the repeated presentations of the same task block may provide more robust estimates of similarity. We also observed that the 2-back vs. 0-back condition contrast explained more variance in spatiotemporal activation patterns ($R^2 = 0.355$) compared to the static activation maps ($R^2 = 0.269$). This may be because the additional information regarding the temporal progression of activation patterns during the task provided more information to distinguish 0-back vs. 2-back conditions. Studies
with a greater sample of tasks are needed to say conclusively that spatiotemporal activation patterns are more discriminative than static activation maps.

Clustering of spatiotemporal activation patterns across all subjects revealed two dominant clusters accounting for 83.94% of the variance among spatiotemporal activation patterns. Visual examination of the centroid spatiotemporal activation patterns of the two clusters revealed overlapping, yet distinct temporal sequences of activation patterns across the duration of each task block. Both were variations on the familiar task-positive and task-negative activation pattern generally observed across task-fMRI experiments (Bolt, Nomi, Yeo, et al., 2017; Duncan, 2010; Hugdahl, Raichle, Mitra, & Specht, 2015). The differences lay in the timing and duration of these patterns across the course of the task block. Cluster one was associated with 0-back conditions, and its spatiotemporal pattern exhibited a diminishing task-positive network activation during the latter half of the task block. Because task-positive activation is often associated with a “focused attention” process (Bolt, Nomi, Yeo, et al., 2017; Hugdahl et al., 2015), this may indicate a waning of attention to the task stimulus during the latter half of the 0-back task block, due to relaxed short-term memory demands. However, task-positive activation was maintained in cluster two spatiotemporal pattern that was associated the 2-back condition. This may indicate a maintained task-driven attention due to the stringent short-term memory demands of the 2-back condition.

Overall, the results of this study suggest that inter-subject variability is pervasive in task-fMRI BOLD activation patterns, both for “static” and temporally extended activation patterns. However, future studies on a larger sample of task paradigms are needed to confirm this conclusion. Categories of a task-fMRI ontology are usually
specified at the population level, without reference to the possibility of inter-subject
variability. The results of this study suggest that caution is required when moving from
inferences regarding categories of task-fMRI ontologies at the population level to the
subject-level. As the results of study two demonstrate, intrinsic differences in BOLD
activation patterns between subjects is likely to a more prevalent factor than
experimentally-controlled manipulations of the task-fMRI environment.
Chapter 4: Summary

In conclusion, results from both studies suggest that a data-driven task-fMRI ontology is a viable project for cognitive neuroscience. While no definitive conclusions can be made regarding a final ontological categorization for task-fMRI, the quantitative framework outlined in these two studies provides a route to that destination. In study one, we observed that a preliminary ontology of four or seven latent cognitive categories provides a simplified description of the observed differences in whole-brain BOLD activation patterns. In study two, we observed that while these categories may provide an adequate description of BOLD activation patterns at the population level, inter-subject variability restrains inferences from these population level distinctions to the subject level. Ultimately, the viability of a data-driven task-fMRI ontology hinges on one’s answer to the question of how fMRI informs cognitive science. Do BOLD recordings from fMRI place any constraint on theories of cognitive processes specified at the behavioral or psychological level? The assumption throughout this dissertation is that BOLD recordings do indeed constrain theorizing at the behavioral or psychological level of analysis. But the viability of this assumption depends on its productivity for cognitive science, and the new hypotheses and theories it generates.
Works Cited


Lynch, C. (2008, September 3). Big data: How do your data grow? [Comments and Opinion]. https://doi.org/10.1038/455028a


Marx, V. (2013, June 12). Biology: The big challenges of big data [News]. https://doi.org/10.1038/498255a


Appendix

A. Brief Discussion of “Activation Maps”

To get a better understanding of these results and how they relate to the cognitive ontology above, it would be a good idea to get a better handle on what the “activation maps” represent and how they are created. Activation maps are created through the application of the mass univariate general linear model (GLM) approach that is implemented in all of the main fMRI software packages: SPM (Friston et al., 1994), FSL (S. M. Smith et al., 2004), and AFNI (Cox, 1996). All packages take the same basic approach: 1) at the subject-level, for each voxel, compute the association between the convolved task design (the hypothesized neural activity convolved with what we expect the BOLD response to look like—the hemodynamic response function) and the BOLD time series of the voxel with a regression model (Figure 6), 2) for the beta coefficient estimate of each voxel, representing the degree of association between the voxel BOLD time series and the convolved task design for that subject, use a random-effects model to compute a group-level estimate across all subjects. This conventional two-step approach, called the two-stages summary statistics approach (Mumford & Poldrack, 2007), is unique to fMRI software packages due to the size of the data; in typical mixed-effects analyses, both levels are estimated simultaneously. The end result is a group-level estimate of the degree of the association between the BOLD time series of each voxel in the brain and the task design, often displayed as a colored brain map.

It is conventional to see the resulting map described as “the increase in BOLD activation from baseline,” as the convolved task design often attempts to model a sustained increase in BOLD activity from the fixation or rest trial/block to the task
trial/block (**Appendix Figure**). This is a bit misleading, because not all time-locked increases in BOLD activity to stimulus/task-block onsets would be captured by a standard convolved task design. For example, the BOLD increase in that voxel could be short-lived or possibly oscillate in response to the task block. Nevertheless, since these estimates often represent a sustained increase from baseline, we refer to them as “activation maps.”

**Subject-Level Regression**

Appendix Figure. **Illustration of Subject-Level GLM (Regression) Approach.** Diagram illustrating the creation of a subject-level activation map from a simple visual checkerboard experiment. For each voxel in the brain, we compute the association between its recorded BOLD signal (the blue time series in the plot below) and the convolved task design (the red line plot or model time series in the plow below)—modelling the block structure of the task (three presentations of a 20s visual checkerboard) with what we expect the BOLD response to look like (the hemodynamic response function). If we do this regression model for every single voxel in the brain, we can represent the beta coefficient values as a colored “activation map”, representing stronger positive associations with lighter colors, and stronger negative associations with cooler colors. Of note, this is a subject-level activation map, the “activation maps” referred to in the text are activation maps computed at the group-level using a two-stage summary statistics approach.
Most task fMRI experimental designs are built on a “subtraction” logic. Designs typically consist of a control condition and an active task condition, where the active condition presumably differs from the control condition in only one cognitive function of interest. One can then subtract the activation map modeled in the control condition from the active condition to examine the activation solely due to the isolated cognitive function.