From brain maps to cognitive ontologies: informatics and the search for mental structure

Russell A. Poldrack[1] and **Tal Yarkoni**[1]

[1]Department of Psychology, Stanford University, Stanford, CA 94305; poldrack@stanford.edu

[2]Department of Psychology, University of Texas at Austin, Austin, TX 78712; tyarkoni@utexas.edu

Abstract

A major goal of cognitive neuroscience is to delineate how brain systems give rise to mental function. Here we review the increasingly large role informatics-driven approaches are playing in such efforts. We begin by reviewing a number of challenges conventional neuroimaging approaches face in trying to delineate brain-cognition mappings—for example, the difficulty in establishing the specificity of postulated associations. Next, we demonstrate how these limitations can potentially be overcome using complementary approaches that emphasize large-scale analysis —including meta-analytic methods that synthesize hundreds or thousands of studies at a time; latent-variable approaches that seek to extract structure from data in a bottom-up manner; and predictive modeling approaches capable of quantitatively inferring mental states from patterns of brain activity. We highlight the underappreciated but critical role for formal cognitive ontologies in helping to clarify, refine, and test theories of brain and cognitive function. Finally, we conclude with a speculative discussion of what future informatics developments may hold for cognitive neuroscience.

Keywords

1. INTRODUCTION

One of the central goals of cognitive neuroscience is to understand how brain systems give rise to cognitive functions, which raises a critical question: What are the cognitive processes that we aim to understand? To grasp the importance of this question, consider an analogy from the field of molecular biology. A central question for this field is how genes are translated into proteins, and answering this question requires a systematic description of the genes and the proteins that are being related. Fortunately for molecular biologists, there are a number of databases that describe all of the proteins and genes that have been discovered across a wide range of species. These databases ground the concepts that researchers in the

domain are studying in a set of objective definitions, allowing different researchers to be sure that they are talking about the same thing. For example, a researcher may describe the object of their study as "DARPP-32", but by providing an accession number in the UniProt database, one could easily ensure that the protein under investigation is the same as one named "Protein phosphatase 1 regulatory subunit 1B" in another paper.

By comparison, cognitive neuroscience is awash in a sea of conflicting terms and concepts. William Uttal summed up this problem clearly in his well-known critique of neuroimaging, "The New Phrenology":

> 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)

In this article we will outline how the field of cognitive neuroscience has begun to address this challenge through the use of tools adapted from the field of biomedical informatics. First we address two fundamental challenges that face the enterprise of cognitive neuroscience. One challenge centers around the difficulty in isolating specific mental functions using psychological tasks. Even if this challenge is solved, a second, deeper problem arises in the establishment of selective mappings between brain systems and mental functions. We argue that the standard approach to neuroimaging is fundamentally unable to deliver such selective mappings. Second, we discuss how large-scale databases enable more powerful analyses to address these challenges. In addition to reviewing conventional benefits of conducting analyses at scale–e.g., aggregating over hundreds or thousands of studies at a time allows estimation of associations with a precision that individual studies typically cannot–we focus on novel inferences that are only possible using such large-scale data. We demonstrate how large-scale databases can help quantify the true specificity of hypothesized structure-function associations by "zooming out" from a single brain circuit or experimental contrast to survey an entire complex landscape of many-to-many mappings between psychological and neural processes. We review data-driven approaches that leverage the scale and breadth of such databases to identify latent components of brain activity and cognitive function. And we illustrate how predictive modeling techniques can combine with large-scale databases to support novel quantitative approaches to the decoding of mental states from brain activity.

Finally, we describe how formal ontologies hold an important key to better describing the structure of the mind and its relation to the brain. We will describe the Cognitive Atlas project, which aims to develop a formal ontology for cognitive neuroscience. We conclude by highlighting some of the future directions that we foresee for the field of cognitive neuroinformatics.

2. INFERENTIAL CHALLENGES FOR NEUROIMAGING

The development and widespread application of modern functional neuroimaging methods such as fMRI has long offered the tantalizing promise that researchers might one day understand how large-scale patterns of brain activity map onto specific mental states or processes. This promise has already been partly realized by the discovery of numerous brain-cognition associations over the past two decades (e.g., the existence of brain regions that preferentially process certain classes of perceptual stimuli, or the increase in activation of a "default mode network" when people are engaged in undirected mental activity Raichle et al. (2001)). However, it has also become increasingly clear that numerous inferential challenges threaten the broad goal of attaining a comprehensive understanding of the joint structure of the mind and brain using functional neuroimaging techniques. Many of these inferential threats are statistical or methodological in nature (e.g., low statistical power, preprocessing and registration problems, etc.), and are outside the scope of this paper. However, a number of threats stem from very basic conceptual challenges that we believe remain widely under-appreciated within the neuroimaging community. Here we discuss two such challenges: first, the difficulty of isolating cognitive functions, and second, the difficulty in establishing specific mappings between brain and behavior.

2.1. Isolating cognitive functions

In principle, identifying the neural substrates of specific cognitive functions using functional neuroimaging would appear to be conceptually straightforward. According to the classical subtraction logic that underlies much of neuroimaging research (Poldrack 2010), it should be possible to identify the neural correlates of specific processes by contrasting experimental conditions that are carefully selected to vary with respect to only a key process of interest. This is sometimes referred to as the assumption of "pure insertion," in reference to the idea (originally attributed to Donders; cf. Sternberg 1969) that one can theoretically add a discrete processing step to an existing task without meaningfully altering the remaining set of processes. For example, by contrasting a condition in which participants passively view visual stimuli with a condition in which participants press a button whenever a new stimulus appears (while holding presentation duration constant across both conditions), one might perhaps be able to cleanly isolate the neural processes associated with planning and executing a motor response.

In practice, of course, things are more complicated. In particular, the logic of cognitive subtraction is notoriously fragile in the face of real-world psychological tasks (e.g. Egeth, Marcus & Bevan 1972; Friston et al. 1996; Jennings et al. 1997). For one thing, the assumption of pure insertion is demonstrably false in many, and perhaps most, cases. Even a simple manipulation of motor responding is unlikely to cleanly isolate motor processes as intended, because the requirement to make a motor response is all but guaranteed to change the way participants deploy attention to the visual stimuli (e.g., it may induce top-down biasing of early visual activity in a proactive effort to identify stimulus changes as soon as they occur—something that would be unlikely during passive viewing). Consistent with this, both behavioral (Egeth, Marcus & Bevan 1972) and neuroimaging (Jennings et al. 1997) studies have shown interactions between task performance (response times and activation

respectively) and response sets. The same fragility is likely to be true for most other experimental approaches as well. For example, parametric designs (in which a single task parameter is varied, such as memory load) rely on a similar "pure modulation" assumption (Poldrack 2010), i.e., that the only change occurring is that of the specific parameter being modulated. Parametric increases in working memory load influence not only WM-related circuits but also motivational and attentional circuits (e.g., as the number of encoded items in a Sternberg task increases, some participants may begin to experience negative emotion due to their inability to perform the task), violating the pure modulation assumption.

Although these problems with the isolation of specific processes using subtractive designs are widely known, it remains common in the neuroimaging literature to conflate experimental manipulations with the specific cognitive functions that are putatively manipulated. The cautions of Cronbach and Meehl (Chronbach & Meehl 1955) against conflating latent constructs with operational measures appear to have been largely forgotten. For example, a search of PubMed reveals more than 1,800 papers whose title or abstract includes the phrase "working memory task." It may not register to most of those authors that in using this term (rather than a more descriptive term such as "Sternberg item recognition task" or "delayed response task"), they are making a theoretical claim, i.e. that the task in question provides a way to isolate a specific mental process called "working memory." This becomes particularly problematic when the mappings between constructs and tasks come into question. For example, there are 99 abstracts in PubMed that include the phrase "N-back working memory task," even though the construct validity of the N-back task as a measure of "working memory" has come into serious question Kane et al. (2007). As we discuss further below, this is a perfect recipe for conceptual confusion.

2.2. Establishing specificity

Even if we suppose, for the sake of argument, that it were possible to employ manipulations in neuroimaging experiments that completely upheld the subtractive assumptions, we would still face the equally daunting problem of establishing the *specificity* of brain-behavior associations. Suppose we found a 100% pure task of working memory that reliably activated lateral prefrontal brain regions when scanned with fMRI. Would we be entitled to conclude that we have established the cognitive function of lateral PFC? No. What such a finding would establish is only that working memory engagement is a *sufficient condition* for activation of LPFC. It would not support the opposite inference—namely, that if LPFC is active, working memory processes must be engaged. The latter inference is invalid because there could in principle be many other psychological processes that also activate LPFC but have little to do with working memory. This difficulty in probabilistically inferring mental function from observed brain activity has been dubbed the problem of *reverse inference* in the cognitive neuroscience literature (Poldrack 2006, 2011). Formally, the reverse inference problem involves inferring the likelihood of engagement of a particular mental process *MP* from a particular activation *A* (which could be a single region or pattern across regions). This can be obtained using Bayes' rule (Poldrack 2006):

$$P\left(MP|A\right) = \frac{P\left(A|MP\right) * P\left(MP\right)}{P\left(A\right)}$$

given some prior $P(MP)$ on the likelihood of the mental process being engaged, and a base rate $P(A)$ for the activation in question. The utility of framing the reverse inference problem in terms of Bayesian inference is that it makes clear that the added value of any activation in identifying the underlying mental process (i.e. the difference between the prior and the posterior probability) is a function of the likelihood of activation in the specific condition *relative to* the base rate of activation in that region. Regions that are more active across all psychological functions will provide less support for any specific reverse inference.

Alternatively, one can describe these inferential challenges in terms of necessary and sufficient conditions. The reason that randomized, controlled experimentation is widely hailed as the gold standard in science is that it can support strong conclusions about causal sufficiency. For example, suppose we experimentally manipulate the nature of the stimuli during a working memory task—presenting, say, faces in one condition and words in another—and observe that the face condition is associated with widespread changes in inferotemporal and frontal brain activity, and also with longer behavioral reaction times. We would be able to definitively conclude that the experimental manipulation we introduced is causally sufficient to produce both the neural and the behavioral changes we observed. However, we cannot conclude the opposite, viz. that engagement of those specific cognitive processes is *necessary* to produce that specific activation or behavioral response.

To see this, consider an example from psychology. Suppose subjects perform a working memory task with two conditions that vary in putative working memory load, and tend to respond more slowly in the high-load condition. It might be natural in such a case to say that increased working memory processing is the *cause* of subjects' slowed responses. But now suppose that the same subjects also perform a visual discrimination task involving two conditions that differ only in the physical size of the on-screen stimuli. Further, suppose that the subjects tend to respond more slowly in the small-stimulus condition. Surely, in this case, few psychologists would want to conclude that slowed responses to smaller visual stimuli must at least partly reflect increased working memory load! What this example illustrates is that claims that seem perfectly reasonable in one context—e.g., that increased reaction time in a WM task reflects increased WM load – often depend critically on tacit background assumptions that are not part of the formal inference. If the context changes, one may then have to invoke a completely different set of putative causes for the same observable outcome. Unless the background conditions thought to differentiate between different contexts can be explicitly modeled, it is not clear how one might quantitatively infer which particular set of causes were responsible for a given outcome.

Relatedly, we also cannot easily draw causal conclusions about the relationship between brain activity changes and behavioral changes. When one observes correlated neural and behavioral changes that make intuitive sense, it can be tempting to interpret the former as the cause of the latter. However, the fact that all behavioral changes must derive from *some* neural changes does not mean that they must derive from the particular neural changes one happens to observe in any given study. For example, the fact that increases in frontal activity are consistently associated with longer reaction times does not entail that the former cause the latter. It is conceivable that frontal increases are the result rather than the cause of variations in reaction time, simply reflecting the fact that participants are processing

information for a longer period of time on trials when they take longer to respond (cf. Yarkoni et al. 2009). Indeed, one under-appreciated implication of the fact that the BOLD signal sums approximately linearly over time (Dale & Buckner 1997) is that any increase in the duration of local processing in a region—*no matter what its cause may be*—is likely to produce a corresponding increase in observed brain activity. Such examples underscore a major challenge to efforts to map the structure of human cognition using neuroimaging, because many of the brain-behavior mappings researchers have drawn in the literature are based largely on observation of concomitant neural and behavioral changes.

3. THE BENEFITS OF LARGE-SCALE INFORMATICS APPROACHES

Importantly, the inferential challenges discussed in the previous section are not intrinsic to cognitive neuroscience or functional neuroimaging, but simply reflect pragmatic constraints on what one can expect to achieve in any single study. The problem of reverse inference, for instance, arises not because it is fundamentally impossible to infer mental states from brain states, but because it is very difficult to contrast a sufficient number of experimental conditions to justify strong claims about the specificity of any individual mapping. It may be feasible to make much more circumscribed claims that are conditioned on specific background conditions (cf. Hutzler 2014; Klein 2012; Machery 2014) — e.g., that, conditional on doing a reading task, some pattern of activity implies orthographic decoding. Similarly, the uncertainty surrounding which cognitive process deserves credit for the effect of a particular experimental task on brain activity is attributable to the impracticality of using dozens of different tasks in every study in order to isolate a specific process by converging operations (Garner, Hake & Eriksen 1956). Fortunately, both of these limitations can be ameliorated by scaling up one's investigation to simultaneously consider the results of *many* different studies. In this section we discuss a number of ways that existing and emerging informatics platforms can help map the structure of human cognition in novel ways and on an unprecedented scale.

3.1. Large-scale meta-analytic structure-to-function mapping

One of the first informatics-driven advances in researchers' ability to map brain-cognition relationships was the development of new statistical methods and software packages for fMRI meta-analysis, and the creation of associated coordinate databases such as BrainMap (Laird, Lancaster & Fox 2005), Brede (Nielsen, Hansen & Balslev 2004), and SumsDB (Dickson, Drury & Van Essen 2001). The BrainMap database, for example, currently contains over 100,000 activation coordinates from over 2,600 fMRI studies that span diverse cognitive domains. All experimental contrasts are annotated with key metadata (e.g., sample size, clinical population(s), etc.) and coded along key dimensions (e.g., stimulus modality, task type, etc.). Drawing on such databases and tools, researchers have conducted hundreds of fMRI meta-analyses on topics ranging from single-word reading (Turkeltaub et al. 2002) to rectal distension in irritable bowel syndrome (Tillisch, Mayer & Labus 2011). By aggregating across dozens, and in some cases hundreds, of fMRI studies, such meta-analyses have been able to overcome the sensitivity limitations associated with many primary fMRI studies (Wager, Lindquist & Kaplan 2007) and provide highly robust estimates of the neural correlates of relatively specific cognitive tasks.

Despite their enormous impact on the field, however, conventional meta-analytic approaches also have important limitations. One is their lack of scalability: because manual annotation and verification of published articles is a time-consuming process, existing coordinate databases such as BrainMap are no longer able to keep up with the growth of the primary literature (cf. Derrfuss & Mar 2009). Thus, as time goes on, it becomes increasingly difficult for investigators to conduct comprehensive meta-analysis of the literature, even in relatively circumscribed domains. A second problem is that conventional meta-analysis approaches, which focus on identifying the brain regions consistently activated by particular cognitive tasks or processes, do not help address the long-standing problem of reverse inference. Consider, for example, the challenge of determining what cognitive function(s) the human anterior insula supports. Individual fMRI studies have implicated this region in the processing of pain (Wager et al. 2004), interoceptive awareness (Critchley et al. 2004), error monitoring (Klein et al. 2007), sustained attention (Dosenbach et al. 2006), phonological processing (Wise et al. 1999), salience (Wiech et al. 2010) and numerous other processes— and this variety is recapitulated in an equally broad range of meta-analyses that also report anterior insula activation, including studies of empathy (Fan et al. 2011), subsequent memory (Kim 2011), and working memory (Owen et al. 2005), among others. While such findings convincingly demonstrate that many different kinds of tasks reliably activate the anterior insula, they provide relatively little insight into what the *specific* function of the anterior insula (or any other region) might be.

In work that addresses both the scalability and inferential limitations of conventional meta-analysis, we recently introduced a novel framework called Neurosynth that supports large-scale synthesis of fMRI data using a fundamentally different approach (Yarkoni et al. 2011). Instead of relying on careful manual annotation of studies, our approach emphasizes automation and scale—effectively trading quality for quantity. Neurosynth uses relatively simple text mining and computational linguistics methods to automatically extract both reported activations and semantic annotations from published articles (see Figure 1). While the data extracted from any individual study are highly susceptible to error, and lack corresponding metadata (e.g., one cannot even reliably determine whether a particular activation represents an increase or decrease in blood flow), the high degree of automation enables Neurosynth to grow in stride with the primary literature at virtually no cost. Consequently, the Neurosynth database has now eclipsed BrainMap in size, with over 10,000 studies and 360,000 discrete activations represented. In keeping with a philosophy of data sharing, reproducibility, and open science, the entire database is made freely available to the community without any use restrictions or requirements of coauthorship.

The relatively comprehensive coverage of the Neurosynth database opens the door to novel kinds of inference—particularly those focused on quantifying the specificity of brain-cognition associations. In an emblematic recent study, Chang and colleagues used Neurosynth to quantitatively "decode" the psychological processes associated with different sectors of the human insula, including the aforementioned anterior sector (Chang et al. 2013). The availability of a relatively comprehensive cross-section of the fMRI literature enabled the authors to quantify not only which kinds of tasks tend to consistently produce anterior insula activity (nearly all of them), but also which processes were *most* likely to

activate the region—thereby providing a measure of specificity that individual fMRI studies or meta-analysis studies focused on experimental comparisons could not (Figure 2). Chang et al. demonstrated that the anterior insula is maximally associated with higher cognitive processes such as task-switching and response inhibition, suggesting that its frequent presence in other kinds of tasks may reflect a fundamental role in basic goal-directed cognitive processes necessary for all kinds of complex cognition (for further discussion, see (Chang et al. 2013)).

To facilitate community adoption of such methods, we have publicly released the entire Neurosynth codebase under a permissive software license, and have developed an interactive web portal (http://neurosynth.org). The Neurosynth website makes it possible, for example, to obtain whole brain reverse inference maps for concepts like reward, episodic memory, or response inhibition, or to generate a rank-ordered list of the psychological concepts most strongly associated with activation at any location in the brain (http://neurosynth.org/locations). Most recently, we have introduced a real-time web interface that supports the kind of open-ended decoding functionality employed in (Chang et al. 2013), enabling other researchers to perform near-instantaneous quantitative reverse inference on uploaded whole-brain statistical maps. Although the results of such analyses have a number of important limitations (discussed below), we believe they nevertheless represent a significant advance over the largely qualitative interpretations that have historically dominated Discussion sections in fMRI articles (cf.(Poldrack 2006)). Notably, they support inferences that are based on interpretation of whole-brain patterns of activity rather than isolated brain regions, and, in virtue of relying on automated analysis of the literature as a whole, they are less susceptible to various cognitive biases known to affect research results (e.g., the tendency to preferentially focus on studies that converge with one's preferred theoretical position).

3.2. Functional parcellation and the search for latent structure

A second, and very different, way of approaching the search for mental structure is to ask whether and how the brain carves the joints of mental function, through an analysis of the latent structure of the neuroimaging data. There is a long history of using statistical methods to try to parcellate the brain into functional units based on correlated activity or connectivity patterns. Early reports that widely-distributed networks of brain regions often modulate in concert (e.g., (Biswal et al. 1995)) quickly led to a diversity of large-scale efforts aimed at extracting a comprehensive set of networks, parcels, or components that capture the dynamics of brain activity—e.g., by using clustering techniques to identify locally-homogeneous sets of voxels (Craddock et al. 2012; Gordon et al. 2014), or matrix factorization methods such as independent components analysis [ICA] (Beckmann & Smith 2004)) to find low-rank approximations to high-dimensional connectivity data (for review, see (Varoquaux & Craddock 2013)). Although such efforts were initially focused solely on the neurobiological level of analysis, researchers soon recognized their potential utility as bridges between brain-based network analysis and cognitive function. In an influential example, Smith et al. (Smith et al. 2009) mapped a set of ICA components extracted from activation coordinates in the BrainMap database onto the domains of tasks used in the associated studies, uncovering a set of plausible mappings between neural networks and

psychological functions. Figure 3 presents a similar analysis performed on the Neurosynth database.

It is also possible to work in the opposite direction, starting with a decomposition of the psychological space and assessing its relationship to brain activity. Poldrack et al. (Poldrack et al. 2012) first performed topic modeling on the text from the Neurosynth database using a technique known as latent Dirichlet allocation (Blei, Ng & Jordan 2003). This method defines a probabilistic generative model for text which involves first sampling from a set of topics, and then sampling words according to their probability under the selected topic. Given a set of documents, the latent topics are inferred using Bayesian estimation, which provides a set of weights for each term and each document in relation to each of the topics. Separate analyses were performed limiting the text to terms related to either psychological functions or brain disorders. These weights were then used to generate maps showing which regions were statistically associated with loading on each topic (referred to as "topic maps"). The topics identified using this technique generated plausible sets of cognitive terms and related maps (e.g, one topic has as its top terms "narrative", "discourse","comprehension", "memory",and "discourse processing", and was associated with activation across the left hemisphere language network; another had top terms "auditory","perception","hearing", "attention", and "listening" and was associated with activation in bilateral auditory cortices). More recently, Yeo and colleagues (Yeo et al. 2014) used a more sophisticated hierarchical topic model to extract a set of "cognitive components" that jointly explain the covariance structure between both cognitive tasks and brain activity—thus attempting to formalize the intuitive idea that the mapping between neural activity in individual brain regions and cognitive tasks may be best described hierarchically (i.e., lower-level units are repeatedly configured into higher-level circuits).

3.3. Decoding mental structure

A third way to approach the challenge of inferring mental structure from neurobiological data is in terms of prediction: Given a neuroimaging dataset, we wish to make the most accurate prediction possible regarding which mental processes are engaged. The coordinate-based approaches described above have been surprisingly successful at making predictions about broad categories, but they generally lack the detailed process-level annotation or the subject-level neuroimaging data necessary to make much more specific predictions. An alternative approach has been to apply machine learning classification and decoding techniques (Haynes & Rees 2006; Norman et al. 2006) to smaller datasets—e.g., to try to determine which of several classes of pictures or words a subject is currently viewing (e.g., (Mitchell et al. 2003; Cox & Savoy 2003)). The earliest work in this domain focused primarily on decoding of specific stimulus or task features within a single individual, but subsequent work has shown that one can decode large-scale psychological functions from fMRI data in a way that generalizes across individuals Mourão-Miranda et al. (2005); Shinkareva et al. (2008); Wager et al. (2013). However, the high decoding accuracies frequently reported in such studies also belie their highly-constrained nature: it is much easier to correctly classify subjects' mental states when there are only a handful of possible candidates (e.g,, discriminating faces versus houses, as opposed to discriminating all possible classes of objects) (cf. Hutzler 2014; Klein 2012; Machery 2014). Thus, the major

open challenge is to combine the respective strengths of these two decoding approaches, with the goal of eventually generating highly accurate, yet relatively unconstrained, predictions about people's mental states.

Although unconstrained decoding of mental states currently lies more in the realm of fiction than of science, there are promising incipient efforts. One recent line of work focuses on reconstructing subjects' experience of a broad range of stimuli under relatively naturalistic conditions—often including identification of previously unseen stimuli. For example, Kay and colleagues use novel encoding models to accurately identify which of 120 natural images subjects were viewing Kay et al. (2008). More recent studies have used similar approaches to reconstruct movie clips from visual cortex activation Nishimoto et al. (2011) or apply semantic labels to thousands of objects and actions Huth et al. (2012). Related work has used large-scale models of semantic structure to decode the identity of words and pictures from brain images Mitchell et al. (2008). Although such studies remain constrained to one particular modality, their use of generative encoding models that can identify entirely new objects represents an important advance over older approaches that discriminate between a small, fixed set of alternatives.

In a different line of work using subject-level data from eight different tasks, we (Poldrack, Halchenko & Hanson 2009) examined whether it was possible to decode which task an individual was engaged in from their brain activity pattern, using a neural network classifier trained on other individuals. We found that it was possible to decode these tasks with greater than 80% accuracy, and that the underlying structure of the trained neural network provided insights into the clusters of tasks that engaged similar neural patterns. In subsequent work using the OpenfMRI database we have shown that it is possible to decode a much larger number of tasks; for example, in (Poldrack et al. 2013) we found that it was possible to obtain greater than 50% accuracy at classifying between 26 different task contrasts from this database. Interestingly, this classification accuracy could be obtained even when the dimensionality was greatly reduced using independent components analysis (reaching asymptote around 100 dimensions), suggesting that it primarily reflects the balance between large-scale neural systems rather than fine-grained patterns of activity. Further analysis of the confusion matrix showed that in many cases similar tasks from different studies were systematically confused by the classifier, suggesting that it was tapping into general cognitive features of those tasks rather than specifics of the particular design.

A more demanding question is whether it is possible to predict the psychological processes underlying the task rather than the task identity. This question has also been addressed in a number of recent studies. Koyejo and Poldrack (Koyejo & Poldrack 2013) used data from the OpenfMRI database that had been annotated manually to specify the putative psychological processes engaged by each of the 26 task contrasts, with the goal of predicting psychological processes rather than task labels. Because many of the contrasts were thought to isolate multiple psychological functions, we used "multilabel" classifiers that have the ability to predict the presence of multiple processes for any particular dataset. We found that it was possible to predict many of the psychological processes with relatively high accuracy, particularly for those processes that occurred relatively often within the database (and thus had more data available for training). In similar work, Schwartz and colleagues (Schwartz,

Thirion & Varoquaux 2013) used the OpenfMRI database to decode task features (such as the stimulus modality and nature of the response), showing similarly strong classification for these features. These studies provide the proof of concept that it should be possible to accurately decode the psychological building blocks of a task from its associated activation patterns.

3.4. Limitations of large-scale, brain-based approaches

The results described above highlight the utility of large-scale, brain-based approaches in discovering interesting relations between psychological functions and brain systems. However, such approaches also have important limitations. Some of these limitations are primarily technical in nature, and reflect current methodological weaknesses that are likely to be overcome in the future via introduction of other novel approaches and informatics platforms. For example, one general problem for virtually all extant meta-analytic databases is the lack of psychologically detailed annotations. This weakness is most prominent in the case of Neurosynth, where data are automatically extracted using relatively simple heuristics that cannot readily identify key metadata fields (e.g., sample size, direction of experimental contrast, etc.); however, even when neuroimaging studies are manually curated, as in the BrainMap database, annotations are largely focused on task-level descriptions rather than underlying psychological processes (e.g., knowing that an *N-back paradigm* with *face stimuli* was used in an experiment does not directly convey whether psychological processes such as active maintenance of information, familiarity detection, phonological rehearsal, and so on, are involved in carrying out the task). We discuss potential solutions to this problem in the next section.

Another technical limitation is that virtually all existing meta-analytic approaches rely on analysis of discrete coordinates reported in published articles rather than continuous whole-brain statistical maps. Coordinate-based meta-analysis (CBMA) is demonstrably inferior to image-based meta-analysis (IBMA), as it unnecessarily discards the vast majority of usable information in the original maps, thereby reducing sensitivity and typically precluding the estimation of continuous effect sizes (Salimi-Khorshidi et al. 2009). In the hopes of facilitating a shift to image-based approaches, we and colleagues are currently working to address this limitation by creating a centralized online repository of whole-brain statistical maps. NeuroVault (http://neurovault.org) is a new platform that allows researchers to quickly upload and annotate their images, facilitating rapid dissemination and interactive visualization of statistical maps, and eventually supporting more powerful meta-analytic syntheses (Gorgolewski et al. 2014).

Finally, there are limitations inherent in fMRI as an imaging methodology. In particular, the temporal resolution of fMRI limits the ability to identify dynamic changes in the millisecond time scale, whereas this is exactly the time scale over which most psychological processes occur. Thus, fMRI maps should be viewed as a composite of all activity occurring during a particular episode of mental activity. To the degree that different psychological functions are distinguished by different *dynamic* combinations of a common set of processing functions on this short time scale, then it may not be possible to disentangle these using fMRI.

In contrast to these purely technical challenges, other limitations of the approaches discussed above are inherent to any effort to approach mind-brain mapping from a purely neurobiological perspective—i.e., by seeking to first identify the "right" functional units at the level of the brain, and then map the revealed structures onto psychological processes. Perhaps the most pressing problem is that a model developed to achieve statistical or theoretical parsimony strictly at a single level of description (e.g., to find the optimal parcellation of functional brain networks given some fixed statistical loss criterion) is not guaranteed to map cleanly onto other levels of description (e.g., cognitive processes). In fact, it almost assuredly will *not*. For instance, it is exceedingly unlikely that there is any single brain region, cluster, or network that corresponds neatly to high-level psychological concepts like episodic recall, working memory, or phonological rehearsal (in much the same way that one would not expect to find a single gene, protein, or neuron type that isomorphically maps onto such high-level concepts). The central question that then arises is what one ought to do in cases where well-established psychological and biological structures do not seem to map well onto one another. For instance, if there is no obvious biological entity that maps cleanly onto the psychological concept of working memory, should we jettison WM from our psychological models in favor of other psychological constructs?

The answer to this question is neither straightforward nor unequivocal. On the one hand, all else being equal, we believe that a model of psychological processes that also maps systematically onto known biological structures is strongly preferable over one that does not —often even when there are other grounds to prefer the latter. For example, suppose that a psychological model with one free parameter captures 92% of the variance in some target behavior while a different model with eight free parameters captures 93%. In such a case, it would seem both theoretically and statistically advisable to favor the simpler model over the more complex, because the additional seven parameters add little incremental value to the behavioral prediction. Yet if the parameters of the more complex model were to each map very cleanly onto well-delineated biological variables, while the single-parameter model correlated very diffusely and non-specifically with brain activity (as it almost certainly must), we would argue that the more complex model is probably more scientifically useful. In this sense, biological discoveries can and should inform the continual revision of psychological theories.

At the same time, we recognize that there may be many cases where there just *isn't* any psychologically tractable model available that simultaneously respects theoretical constraints from both psychology and biology. For example, there is no guarantee that there is *any* viable replacement for the concept of "working memory" that would both (a) map cleanly onto underlying biological structures and (b) remain sufficiently compact and psychologically interpretable to be useful in practice. Would it be advantageous to eliminate a high-level term like "working memory" from our scientific lexicon if the only way to cover approximately the same territory with a biologically-detailed model is to introduce a large disjunctive set of of separate mechanisms? This question largely echoes earlier criticisms of reductionism (Fodor 1974) —i.e., the mere fact that a "fuzzy" higher-level description can *in principle* be replaced by a lower-level description does not mean that the lower-level description will necessarily be more useful in practice. Ultimately the question

will rest on whether such a new framework is more scientifically productive than the current framework, which is unanswerable until the new framework is proposed and tested.

Unfortunately, we know of no algorithmic way to distinguish cases where a psychological concept has outlived its utility from those where a concept simply lacks any viable biologically-inspired replacement but remains useful. In practice, this is a problem that researchers may always have to navigate on a case-by-case basis—and disagreements between researchers will certainly arise in the process. We would argue strongly, however, that informatics-driven methodologies can make it much easier to navigate this problem. In the next section, we discuss ongoing efforts to develop formal cognitive ontologies that can help clarify conceptual definitions, distinguish genuine substantive disagreements from pointless terminological disputes, and generate new ways of studying and thinking about causal relationships within and between different levels of description.

4. TOWARDS A COGNITIVE ONTOLOGY

4.1. The need for formal representations

We have argued above that the question "what are the parts of the mind?" cannot be successfully addressed strictly through a bottom-up perspective that first asks "what are the functional units of the brain?" and only then seeks to map the resulting components onto the psychological space (cf. Price & Friston 2005). But we have also suggested that simply inverting this process and adopting a purely top-down approach—i.e., taking well-established psychological constructs such as working memory and seeking their underlying neural substrates—is not likely to prove much more fruitful. Aside from the methodological and conceptual problems with such an approach discussed in the previous sections (e.g., the problem of reverse inference), there is an arguably even more fundamental problem, which is that psychologists rarely agree on the meaning of the constructs under investigation. For example, the term "working memory" has been defined as:

- "the manipulation and use of information [in short-term memory] to guide behavior." Larocque, Lewis-Peacock & Postle (2014)

- "working memory involves the process of active maintenance of a limited amount of information" Jeneson & Squire (2012)

- "a hypothetical cognitive system responsible for providing access to information required for ongoing cognitive processes" Wilhelm, Hildebrandt & Oberauer (2013)

- "Working memory subsumes the capability to memorize, retrieve and utilize information for a limited period of time" Rottschy et al. (2012)

This diversity of views may reflect what Walter Mischel has called the "toothbrush problem" in psychology: "Psychologists treat other peoples' theories like toothbrushes — no self-respecting person wants to use anyone else's" (Mischel 2008). With such divergent definitions of constructs (and equally divergent tasks used to measure them), how can we expect to find consistent mappings between mental constructs and brain systems?

Of course, it is hardly surprising that disagreements should arise over how to delineate and describe an organ as complex as the human mind. And there is nothing intrinsically wrong with having a diversity of opinions. The concern, however, is that psychologists do not seem to have well-established procedures for effectively resolving such differences. For instance, how should we determine whether working memory is best defined in terms of maintenance and manipulation of information in a short-term memory buffer, or in terms of the ability to flexibly recruit other cognitive resources in support of current goals? Should the label 'working memory' apply solely to a central executive mechanism, or also encompass slave systems such as the phonological loop and visuospatial sketchpad, as in Baddeley's influential model (Baddeley 1992)?. The textbook approach to such disputes is to devise a "critical experiment" that can offer definitive evidence in favor of one theory over another. But in practice, it is exceedingly difficult to identify real-world cases where a critical experiment has actually prompted the abandonment of a theory (cf. (Greenwald 2012)).

We suggest that much of the difficulty resolving theoretical differences is due to the informal nature of most theoretical claims. At present, there is no unifying framework that allows researchers to represent their theories and definitions in a structured, formal way; while formalization will not resolve differences on its own, it makes differences clearer and thus more amenable to testing. Until recently there was no resource we know of that allowed one to easily determine which cognitive processes the Sternberg task invokes, what clinical conditions are associated with impairments of grammar learning, or what set of mental states can be considered instances of emotion. We believe the development of a formal framework for specifying relationships between psychological concepts and tasks would substantially advance our ability to map the structure of human cognition and its underlying neurobiological bases. Notably, there is considerable precedent for such a development in other biomedical fields, where formal ontologies have played a critical role in facilitating hypothesis testing and scientific exploration alike.

4.2. Formal ontologies as a potential solution

One of the most basic questions that is raised by cognitive neuroscience (and that one might reasonably expect psychologists to be able to answer) is: "What are the parts of the mind?" If one had asked this question of a psychologist in the eighteenth century, an answer would have been quickly forthcoming in the form of a list of mental faculties, which were adopted by the earliest "brain mapping" researchers (i.e. phrenologists) as the basis for their structure-function mapping. However, today virtually no psychologists would have an answer to this question. The closest that one might come would be to examine the index of a cognitive psychology textbook, but one would quickly find that there is no systematic description of how psychologists currently characterize the structure of the mind. Given that the goal of cognitive neuroscience is to map mental functions onto brain systems, this poses a fundamental problem.

Compare this situation to a similar question in biology: "What are all of the biological functions that occur within a cell?" As recently as the 1990's, one would have been similarly challenged to find a systematic answer to this question. However, in the late 1990's a group of biologists and informatics experts developed a consortium (called the Gene Ontology

Consortium) that began to develop a formal knowledge base (or *ontology*) to represent the state of current knowledge regarding the structure and function of biological systems Ashburner et al. (2000). Today, one can visit the Gene Ontology web site (http://www.geneonotology.org) and obtain a comprehensive formal description of cellular components, biological processes, and molecular functions.

4.3. What is an ontology?

The term "ontology" is used here to refer to an "explicit specification of a conceptualization" Gruber (1993), or more generally as a formal description of a knowledge structure; this usage of the term, which arises from computer science, is related but distinct from the usage of the term in philosophy, where it often refers to the entities that are postulated by a particular theory Quine (1948) or more generally to the nature of existence. At its base, a formal ontology specifies the entities that exist within a domain along with the relations between those entities Bard & Rhee (2004). For example, within the Gene Ontology, the entity "rough endoplasmic reticulum" has the relation *is-a* to the entity "endoplasmic reticulum" (Figure 4), meaning that the former is agreed upon to be a particular kind of the latter.

The Gene Ontology currently has entries for more than 26,000 biological processes, almost 10,000 molecular functions, and more than 3600 cellular components (and it is only one of many ontologies that have been developed within the bioinformatics community to describe various levels of biological function and structure). Given that each of these entries was manually curated, this represents a massive investment of human time. Why would researchers put so much time and effort into doing this? We will review several important benefits (for others, see Bard & Rhee (2004); Bodenreider & Stevens (2006); Rubin, Shah & Noy (2008)).

Controlled vocabulary with unique identifiers—The establishment of an agreed-upon ontology provides researchers in the field with a controlled vocabulary for description of biological entities, each of which has a unique identifier (e.g., the identifier for "endoplasmic reticulum" is GO:0005789). This provides a machine-readable way to describe each entity, which can remain consistent even if the field decides to change the name of the entity. It also provides a means to link between different databases, which has proven remarkably powerful in biomedical informatics (Bard & Rhee 2004).

Framework for annotation—One of the most important aspects of the ontology is that it provides a framework through which to relate the ontology entities to other types of data, which in the context of bioinformatics is referred to as *annotation*. Gene Ontology entities are annotated by associating them with specific gene products (proteins or RNAs). The Gene Ontology annotation database currently has over two million annotation entries, which reflect many different types of evidence relating specific entities to specific gene products. For example, a search for annotations of the entity "endoplasmic reticulum" identified 119,040 entries involving 74,239 proteins across all species. The availability of a database of these annotations has enabled a number of new tools for understanding biological data. Most important, it has provided the ability to assess, for any specific set of genes, which

entities in the ontology are likely to be "enriched" for that set of genes, thus providing insights into the larger biological context for the experiment (Rhee et al. 2008).

Inferring relations—Ontologies are generally specified using formal knowledge representation systems, or ontology languages, such as the Ontology Web Language (OWL). Once specified this way, formal reasoning systems can be used to infer relations between entities that are not explicitly specified in the ontology. For example, if we know that "rough endoplasmic reticulum" is a kind of "endoplasmic reticulum", and that "endoplasmic reticulum" is a kind of "cytoplasmic part", we can infer that "rough endoplasmic reticulum" is a kind of "cytoplasmic part".

4.4. A pragmatic approach to ontology building

An important consideration when developing a formal ontology of mental processes is that psychological processes are inherently "fuzzier", or less structured, than biological processes. In many biomedical domains, the basic units of analysis correspond to well-defined physical entities, and are not up for serious debate. For example, it's highly unlikely that ten or twenty years hence, geneticists will decide that they were wrong all along about the nucleotide bases that make up the human genome, and will move to eliminate all talk of cytosine, guanine, adenine, and thymine from the lexicon. In contrast, major psychological concepts appear to fall in and out of favor with some regularity—typically without having been demonstrably refuted by any critical experiment (Greenwald 2012). More generally, it is not clear that a question like what are the fundamental building blocks of human cognition? admits of a clear-cut answer in the same way that one can unambiguously identify the letters of the human genome. Is working memory a more basic concept than executive control or cognitive control? Does it make sense to speak of perception as a basic concept, or is that a purely extensional definition that is best ignored in favor of individual sensory systems like vision and audition? Will concepts like love and hate find their place in a formal ontology of the mind, or are they merely folk psychological abstractions to be abolished as science progresses, in the way that some philosophers once envisioned (Churchland 1981)?

The critical point here is not just that there is presently no consensus on such questions; it is that they very likely admit of no single right answer. There is little reason to suppose that the extremely complex and high-dimensional structure of human cognition can be neatly reduced to a much lower-dimensional, "human-readable" description without substantial loss of fidelity. Nor is it clear what criteria one could use to unambiguously distinguish between good and bad models. Should researchers privilege theoretical parsimony, such that a good model is one that maps well onto theoretical entities identified by prior scientific investigation (e.g., nodes in a cognitive ontology should strive to attain a one-toone mapping with neurobiological structures)? Or should they favor statistical parsimony, such that if two models explain the same amount of behavioral variance, the simpler one is to be preferred, even if the more complex one maps more sensibly onto underlying biological entities? The answer will undoubtedly depend on individual researchers' goals and preferences.

Importantly, however, the principled absence of a single unassailable description of the structure of human cognition does not diminish the need for a formal ontology of psychological processes. If anything, the contrary is true. Many of the theoretical disputes that arise in psychology are, we submit, driven to a large extent by tacit differences in terminology that ramify as substantive disagreements. For example, in the personality literature, researchers have long debated whether the "fundamental feature" of Extraversion is reward sensitivity, positive affect, or sociability (Ashton, Lee & Paunonen 2002; Lucas & Diener 2001; Lucas et al. 2000; Lucas, Le & Dyrenforth 2008; Smillie et al. 2012). In our view, it is not clear that there is a definitive answer to this question. It may well be that different research communities are simply applying the label Extraversion to different (though partly overlapping) behaviors—in which case there would be no more utility in trying to determine what the 'true' definition of Extraversion is than in trying to arbitrate between two formal color standards that disagree as to whether aquamarine is a shade of blue or of green. The ability to formally clarify and translate between different lexica would thus be a major boon to theoreticians.

From a purely pragmatic standpoint, one can think of a good ontology as a kind of universal language that dramatically reduces the likelihood of miscommunication between researchers by enabling statements to be defined in more formal and less ambiguous terms—even when the mapping between terms and their referents is still not perfect. At their best, good ontologies can serve as sophisticated inference engines capable of informing and even answering certain kinds of questions that the unaided human mind is known to have trouble with—for example, by identifying when two terms are being used interchangeably; when a single term is being used in multiple conflicting ways; when a relationship between two concepts is likely to exist even though one has not yet been reported; when the putative nomological network (Chronbach & Meehl 1955) of a given concept contains internal inconsistencies; and so on.

4.5. The Cognitive Atlas

The success and broad utility of biomedical ontologies such as the Gene Ontology inspired one of us (RP) to ask whether it would be possible to address these problems by developing a formal ontology of mental processes and tasks, which led to the establishment of the Cognitive Atlas (http://www.cognitiveatlas.org/ Poldrack et al. (2011). The broad goal of the Cognitive Atlas is to serve as an open collaborative knowledge base for psychological science. It is meant to capture two primary forms of knowledge. First, it aims to define psychological constructs in order to provide consensus definitions that can serve as the basis for accurate scientific communication and discussion. A fundamental distinction made within the Cognitive Atlas is between *mental concepts*, which refer to putative but unobservable psychological processes or structures, and *mental tasks*, which are the objective operations used to measure those putative constructs (see Figure 5). This distinction follows the previously noted admonition by Chronbach and Meehl (Chronbach & Meehl 1955) regarding the separation of latent constructs and the operations used to measure them. Second, the project aims to establish a knowledge base of the relations within and between mental tasks and mental concepts. In a sense, it is these relations that form a major part of the basis for psychological theories. For example, Baddeley's theory of working

memory could be specified in terms of a set of concepts (e.g., "phonological loop", "acoustic store", and "central executive"), a set of relations between these concepts (e.g., "acoustic store is part of phonological loop"), and relations to tasks (e.g, "the acoustic store is measured by the comparison of acoustically similar versus dissimilar words on a short-term memory task").

Within the Cognitive Atlas, mental tasks are described in terms of three primary features: *conditions* (which specify different conditions of measurement), *contrasts* (which specify either comparisons between conditions or relationships with continuous variables), and *indicators* (which specify variables that are measured within the task; these could reflect behavioral, neural, or other physiological measurements). In order to capture the relations between tasks and concepts, we defined a novel ontological relationship (measured-by) that denotes the fact that a specific concept is measured by a specific task. Importantly, concepts are not related to the overall representation of a task, but rather to specific contrasts. This design decision recognizes the subtractive/contrastive logic of psychological measurement: researchers are typically interested in comparisons between conditions that vary in some specific set of putative mental processes, rather than in the value of a specific indicator. For example, the concept of the phonological loop might be measured by the contrast of phonologically similar versus dissimilar items on the Sternberg delayed recognition task. Other contrasts within the task could be related to other mental concepts; e.g., the contrast between target-present versus target-absent probe trials might be thought to reflect some aspect of decisional processes, rather than measuring working memory.

The description of tasks within the Cognitive Atlas is relatively abstract, and does not describe specific aspects of the stimuli, responses, or instructions for the task. A parallel effort, known as the Cognitive Paradigm Ontology (or CogPO) (Turner & Laird 2012), is developing the means to describe tasks in more detail. Ultimately it should be possible to link these two ontologies, in order to span directly from psychological processes outlined in the Cognitive Atlas to detailed operational descriptions of tasks. This would allow the proper annotation of cases where specific task implementation details are critical determinants of the psychological processes that are engaged (e.g., changes in the architecture of task switching in relation to the amount of time available for preparation prior to switching (Rogers & Monsell 1995)).

5. FUTURE DEVELOPMENTS

The projects outlined above represent the first steps in the development of a cognitive neuroinformatics that can support strong inferences regarding the relation between brain systems and psychological functions. What new advances will the next few years bring? In a short review several years ago (Yarkoni et al. 2010), we briefly considered the question: What will cognitive neuroscience look like 10 years from now? Some of the developments we anticipated at the time included: fully automated quantitative mapping between cognitive and neural states; intelligent preprocessing and analysis pipelines that evaluate local data in relation to global databases; integration of neuroimaging databases with other kinds of data —e.g., functional genomic repositories; introduction of centralized neuroimaging data repositories; and integration of formal ontologies and formal method descriptions into fMRI

analysis software. Now, five years on (and halfway through our earlier forecast horizon), we are more optimistic than ever about the prospects for a cumulative, integrative, informatics-driven science of the human mind/brain. A number of the developments we anticipated five years ago already have realized implementations that we discuss above (e.g., centralized data repositories like OpenFMRI and NeuroVault, and the ability to instantly decode maps uploaded to NeuroVault using Neurosynth). Others are in very early stages of development. And then, of course, there are other important emerging projects that we did not anticipate at all five years ago. Here we outline a few recent developments of particular interest.

Towards open, standardized, and centralized data sharing

The benefits of cognitive ontologies to scientific discovery become most apparent when applied at scale. A critical component of ongoing and future efforts will therefore be the centralized aggregation and organization of neuroimaging and psychology data. Currently, major ongoing efforts are directed at developing machine-readable standards for representing neuroimaging data (e.g., the Neuroimaging Data Model; http://nidm.nidash.org/); creating reproducible, shareable, open-source analysis pipelines (e.g., the Nipype framework; (Gorgolewski et al. 2011)); and the establishment of open resources for the sharing of both raw fMRI datasets (e.g., OpenFMRI; (Poldrack et al. 2013)) and statistical images (e.g., NeuroVault; http://www.neurovault.org). Some work has started to combine these datasets with ontologies of psychological processes (e.g., (Poldrack et al. 2012)), but the rapid development and growth of these databases will likely enable much more powerful analyses in the future.

Crowdsourced annotation

Despite the recent successes of automated meta-analysis approaches discussed above, it is clear that careful human consideration and annotation of neuroimaging data remains a critical component of most investigations, and is unlikely to be replaced by machine learning approaches in the near future. We suggest that the next wave of advances in the area of neuroimaging meta-analysis may result from successful hybridization of manual and automated approaches—and particularly from development of user-friendly crowdsourcing interfaces that allow researchers to easily apply their expertise to manual curation of communal databases. A promising prototype is BrainSpell (http://brainspell.org), a website that allows users to manually validate, annotate, and tag all data presently in Neurosynth—potentially providing all of the benefits of manual curation for substantially less effort than full manual entry would require. An ongoing challenge, however, is to develop effective incentives for participation in such efforts. One largely unexplored approach in this area is the kind of "gamification" successfully achieved in other domains using platforms such as FoldIt (Khatib et al. 2011) and EyeWire (http://eyewire.org).

Using ontologies to resolve psychological debates

One of the greatest promises of a comprehensive formal ontology of cognitive processes lies in the potential to develop a kind of formal inferential engine that enables researchers to compute well-defined operations over its nodes and relationships, thereby informing, and in some cases even resolving, ongoing theoretical debates. For example, within the Cognitive Atlas we define a measured-by relationship that indicates that a given task contrast (e.g., the

high versus low load conditions on the Sternberg item recognition task) depends critically on a particular psychological concept (e.g., working memory). In the future, we could define a comparison operation that takes two concepts as input and returns separate lists of all known task contrasts that (a) tap both constructs, and (b) tap only one of the constructs. Further, we could define additional operations like similarity or difference that take two lists of contrasts (or concepts) and return either quantitative metrics of similarity (e.g., based on computing the similarity of two nodes' local neighborhood or network structure parameters), or a list of concepts (or tasks) that maximally distinguishes the two inputs. Given such a platform, it could conceivably turn out, for example, that whether the episodic buffer is or is not a central part of working memory is largely a definitional matter: two researchers might each feed in a list of what they consider to be critical working memory tasks, only to find that they are talking past each other, inasmuch as the disagreement is purely extensional (i.e., there is no implied contradiction in which other latent concepts 'working memory' is linked to, but only in how broadly the label is applied to individual tasks).

Similarly, in cases where there are genuine substantive disagreements between theories, a formal ontology of cognition could help focus attention appropriately. For example, there is debate in the executive function literature over whether performance on tasks requiring suppression of a prepotent or ongoing response in favor of a different response requires an active inhibitory process or whether it instead can rely solely on competitive inhibition within a network (Munakata et al. 2011; Aron, Robbins & Poldrack 2004). Each of these theories makes different predictions regarding the similarity (i.e. covariance) of both behavioral and neural activation patterns across a large number of tasks. Given an appropriate annotation of a sufficiently large dataset, one could directly assess which of these theories provides a better fit to the observed data (e.g., using analogs to structural equation modeling), and also potentially demonstrate which specific set of concepts needs to be experimentally compared in order to most powerfully assess the specific theoretical debate.

A unified, interoperable ecosystem

Perhaps the most promising development of the coming years will be the increasing convergence and interoperability between diverse resources for informatics-driven investigation of the human mind/brain. We anticipate the relatively near-term emergence of a unified, interoperable ecosystem made up of dozens of individual services that all loosely follow the same standards and protocols, enabling researchers to construct automated pipelines that easily integrate currently disparate resources. Optimistically, we predict that within a few years, researchers will be able to easily (i.e., without requiring advanced technical skills) upload raw data they have acquired and annotated to centralized platforms that run state-of-the-art cloud-based processing and analysis pipelines; interactively explore the results of such analyses via rich, user-friendly web interfaces that include extensive literature-based quantitative interpretation and allow easy piping to other third-party services; and use ontology-driven inference engines to conduct sophisticated, highly customized meta-analyses that draw on thousands of datasets acquired and deposited using similar platforms.

6. CONCLUSION

The field of cognitive neuroscience faces a number of daunting challenges in its attempt to understand the relation between brain systems and psychological functions. We have argued that the most commonly used approaches in cognitive neuroscience are fundamentally unable to identify the kind of selective associations between neural structure and psychological function that are the presumed goal of the field, but that this question can be profitably addressed using informatics approaches that employ large-scale databases and formal ontologies. We predict that such approaches will become increasingly common in psychology as they have in biology, and that this will provide a new pathway towards discoveries regarding how neural computations give rise to mental life.

ACKNOWLEDGMENTS

GLOSSARY

Annotation	The description of relationships between a dataset and other concepts or datasets.
Classifier	A statistical tool used to classify a new observation into one of a set of discrete classes.
Decoding	The use of neuroimaging data to classify mental activity.
Ontology	A formal description of the concepts assumed to exist within a particular domain, and their relationships.

LITERATURE CITED

Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex. Trends Cogn. Sci. 2004; 8:170–7. [PubMed: 15050513]

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. Gene ontology: tool for the unification of biology. the gene ontology consortium. Nat. Genet. 2000; 25:25–29. [PubMed: 10802651]

Ashton MC, Lee K, Paunonen SV. What is the central feature of extraversion? social attention versus reward sensitivity. J. Pers. Soc. Psychol. 2002; 83:245–52. [PubMed: 12088129]

Baddeley A. Working memory. Science. 1992; 255:556–9. [PubMed: 1736359]

Bard JBL, Rhee SY. Ontologies in biology: design, applications and future challenges. Nat. Rev. Genet. 2004; 5:213–222. [PubMed: 14970823]

Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging. 2004; 23:137–52. [PubMed: 14964560]

Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magn. Reson. Med. 1995; 34:537–41. [PubMed: 8524021]

Blei D, Ng A, Jordan M. Latent dirichlet allocation. J. Mach. Learn. Res. 2003; 3:993–1022.

Bodenreider O, Stevens R. Bio-ontologies: current trends and future directions. Brief. Bioinform. 2006; 7:256–74. [PubMed: 16899495]

Chang LJ, Yarkoni T, Khaw MW, Sanfey AG. Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. Cereb. Cortex. 2013; 23:739–49. [PubMed: 22437053]

Chronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol. Bull. 1955; 52:281–302. [PubMed: 13245896]

Churchland PM. Eliminative materialism and the propositional attitudes. J. Phil. 1981; 78:67–90.

Cox DD, Savoy RL. Functional magnetic resonance imaging (fmri) "brain reading": detecting and classifying distributed patterns of fmri activity in human visual cortex. Neuroimage. 2003; 19:261–70. [PubMed: 12814577]

Craddock RC, James GA, Holtzheimer PE 3rd, Hu XP, Mayberg HS. A whole brain fmri atlas generated via spatially constrained spectral clustering. Hum. Brain. Mapp. 2012; 33:1914–28. [PubMed: 21769991]

Critchley HD, Wiens S, Rotshtein P, Ohman A, Dolan RJ. Neural systems supporting interoceptive awareness. Nat. Neurosci. 2004; 7:189–95. [PubMed: 14730305]

Dale AM, Buckner RL. Selective averaging of rapidly presented individual trials using fmri. Hum. Brain. Mapp. 1997; 5:329–40. [PubMed: 20408237]

Derrfuss J, Mar RA. Lost in localization: the need for a universal coordinate database. Neuroimage. 2009; 48:1–7. [PubMed: 19457374]

Dickson J, Drury H, Van Essen DC. 'the surface management system' (sums) database: a surface-based database to aid cortical surface reconstruction, visualization and analysis. Philos. Trans. R. Soc. Lond. B Biol. Sci. 2001; 356:1277–92. [PubMed: 11545703]

Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, et al. A core system for the implementation of task sets. Neuron. 2006; 50:799–812. [PubMed: 16731517]

Egeth H, Marcus N, Bevan W. Target-set and response-set interaction: implications for models of human information processing. Science. 1972; 176:1447–8. [PubMed: 17834655]

Fan Y, Duncan NW, de Greck M, Northoff G. Is there a core neural network in empathy? an fmri based quantitative meta-analysis. Neurosci. Biobehav. Rev. 2011; 35:903–11. [PubMed: 20974173]

Fodor JA. Special sciences (or: The disunity of science as a working hypothesis). Synthese. 1974; 28:97–115.

Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RS, Dolan RJ. The trouble with cognitive subtraction. Neuroimage. 1996; 4:97–104. [PubMed: 9345501]

Garner WR, Hake HW, Eriksen CW. Operationism and the concept of perception. Psychol. Rev. 1956; 63:149–59. [PubMed: 13323170]

Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. Generation and evaluation of a cortical area parcellation from resting-state correlations. Cereb. Cortex. 2014

Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front. Neuroinform. 2011; 5:13. [PubMed: 21897815]

Gorgolewski KJ, Varoquaux G, Rivera G, Schwartz Y, Ghosh SS, et al. Neurovault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. bioRxiv. 2014

Greenwald AG. There Is Nothing So Theoretical as a Good Method. Persp. Psychol. Sci. 2012; 7:99–108.

Gruber T. A translation approach to portable ontology specifications. Knowledge Acquisition. 1993; 5:199–220.

Haynes JD, Rees G. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 2006; 7:523–34. [PubMed: 16791142]

Huth AG, Nishimoto S, Vu AT, Gallant JL. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron. 2012; 76:1210–24. [PubMed: 23259955]

Hutzler F. Reverse inference is not a fallacy per se: cognitive processes can be inferred from functional imaging data. Neuroimage. 2014; 84:1061–9. [PubMed: 23313571]

Jeneson A, Squire LR. Working memory, long-term memory, and medial temporal lobe function. Learn. Mem. 2012; 19:15–25. [PubMed: 22180053]

Jennings JM, McIntosh AR, Kapur S, Tulving E, Houle S. Cognitive subtractions may not add up: the interaction between semantic processing and response mode. Neuroimage. 1997; 5:229–39. [PubMed: 9345552]

Kane MJ, Conway ARA, Miura TK, Colflesh GJH. Working memory, attention control, and the n-back task: a question of construct validity. J. Exp. Psychol. Learn. Mem. Cogn. 2007; 33:615–22. [PubMed: 17470009]

Kay KN, Naselaris T, Prenger RJ, Gallant JL. Identifying natural images from human brain activity. Nature. 2008; 452:352–5. [PubMed: 18322462]

Khatib F, Cooper S, Tyka MD, Xu K, Makedon I, et al. Algorithm discovery by protein folding game players. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:18949–53. [PubMed: 22065763]

Kim H. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fmri studies. Neuroimage. 2011; 54:2446–61. [PubMed: 20869446]

Klein C. Cognitive ontology and region- versus network-oriented analyses. Phil. Sci. 2012; 79:952–960.

Klein TA, Endrass T, Kathmann N, Neumann J, von Cramon DY, Ullsperger M. Neural correlates of error awareness. Neuroimage. 2007; 34:1774–81. [PubMed: 17185003]

Koyejo O, Poldrack RA. NIPS Workshop on Machine Learning and Interpretation in. Neuroimaging. 2013

Laird AR, Lancaster JL, Fox PT. Brainmap: the social evolution of a human brain mapping database. Neuroinformatics. 2005; 3:65–78. [PubMed: 15897617]

Larocque JJ, Lewis-Peacock JA, Postle BR. Multiple neural states of representation in short-term memory? it's a matter of attention. Front. Hum. Neurosci. 2014; 8:5. [PubMed: 24478671]

Lucas RE, Diener E. Understanding extraverts' enjoyment of social situations: the importance of pleasantness. J. Pers. Soc. Psychol. 2001; 81:343–56. [PubMed: 11519937]

Lucas RE, Diener E, Grob A, Suh EM, Shao L. Cross-cultural evidence for the fundamental features of extraversion. J. Pers. Soc. Psychol. 2000; 79:452–68. [PubMed: 10981846]

Lucas RE, Le K, Dyrenforth PS. Explaining the extraversion/positive affect relation: sociability cannot account for extraverts' greater happiness. J. Pers. 2008; 76:385–414. [PubMed: 18399958]

Machery E. In defense of reverse inference. Brit. J. Phil. Sci. 2014; 65:251–267.

Mischel W. The toothbrush problem. APS Observer. 2008:11.

Mitchell TM, Hutchinson R, Just MA, Niculescu RS, Pereira F, Wang X. Classifying instantaneous cognitive states from fmri data. AMIA Annu. Symp. Proc. 2003; 465:9.

Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, et al. Predicting human brain activity associated with the meanings of nouns. Science. 2008; 320:1191–5. [PubMed: 18511683]

Mourão-Miranda J, Bokde ALW, Born C, Hampel H, Stetter M. Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional mri data. Neuroimage. 2005; 28:980–95. [PubMed: 16275139]

Munakata Y, Herd SA, Chatham CH, Depue BE, Banich MT, O'Reilly RC. A unified framework for inhibitory control. Trends Cogn. Sci. 2011; 15:453–9. [PubMed: 21889391]

Nielsen FA, Hansen LK, Balslev D. Mining for associations between text and brain activation in a functional neuroimaging database. Neuroinformatics. 2004; 2:369–80. [PubMed: 15800369]

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL. Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 2011; 21:1641–6. [PubMed: 21945275]

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fmri data. Trends Cogn. Sci. 2006; 10:424–30. [PubMed: 16899397]

Owen AM, McMillan KM, Laird AR, Bullmore E. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum. Brain. Mapp. 2005; 25:46–59. [PubMed: 15846822]

Poldrack RA. Can cognitive processes be inferred from neuroimaging data? Trends Cogn. Sci. 2006; 10:59–63. [PubMed: 16406760]

Poldrack, RA. Foundational Issues in Human Brain Mapping. Hanson, SJ.; Bunzl, M., editors. MIT Press; Cambridge, MA: 2010. p. 147-160.

Poldrack RA. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. Neuron. 2011; 72:692–7. [PubMed: 22153367]

Poldrack RA, Barch DM, Mitchell JP, Wager TD, Wagner AD, et al. Toward open sharing of task-based fmri data: the openfmri project. Front. Neuroinform. 2013; 7:12. [PubMed: 23847528]

Poldrack RA, Halchenko YO, Hanson SJ. Decoding the large-scale structure of brain function by classifying mental states across individuals. Psychol. Sci. 2009

Poldrack RA, Kittur A, Kalar D, Miller E, Seppa C, et al. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. Front. Neuroinform. 2011; 5:17. [PubMed: 21922006]

Poldrack RA, Mumford JA, Schonberg T, Kalar D, Barman B, Yarkoni T. Discovering relations between mind, brain, and mental disorders using topic mapping. PLoS Comput. Biol. 2012; 8:e1002707. [PubMed: 23071428]

Price C, Friston K. Functional ontologies for cognition: The systematic definition of structure and function. Cogn. Neuropsychol. 2005; 22:262–275. [PubMed: 21038249]

Quine WVO. On what there is. Rev. Metaphysics. 1948

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. 2001; 98:676–82. [PubMed: 11209064]

Rhee SY, Wood V, Dolinski K, Draghici S. Use and misuse of the gene ontology annotations. Nat. Rev. Genet. 2008; 9:509–15. [PubMed: 18475267]

Rogers RD, Monsell S. Costs of a predictable switch between simple cognitive tasks. J. Exp. Psychol.: Gen. 1995; 124:207–231.

Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, et al. Modelling neural correlates of working memory: a coordinate-based meta-analysis. Neuroimage. 2012; 60:830–46. [PubMed: 22178808]

Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. Brief. Bioinform. 2008; 9:75–90. [PubMed: 18077472]

Salimi-Khorshidi G, Smith SM, Keltner JR, Wager TD, Nichols TE. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. Neuroimage. 2009; 45:810–23. [PubMed: 19166944]

Schwartz, Y.; Thirion, B.; Varoquaux, G. Advances in Neural Information Processing Systems 26. Burges, C.; Bottou, L.; Welling, M.; Ghahramani, Z.; Weinberger, K., editors. Curran Associates, Inc.; 2013. p. 1673-1681.

Shinkareva SV, Mason RA, Malave VL, Wang W, Mitchell TM, Just MA. Using fmri brain activation to identify cognitive states associated with perception of tools and dwellings. PLoS One. 2008; 3:e1394. [PubMed: 18167553]

Smillie LD, Cooper AJ, Wilt J, Revelle W. Do extraverts get more bang for the buck? refining the affective-reactivity hypothesis of extraversion. J. Pers. Soc. Psychol. 2012; 103:306–26. [PubMed: 22582899]

Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, et al. Correspondence of the brain's functional architecture during activation and rest. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:13040–5. [PubMed: 19620724]

Sternberg S. Memory-scanning: mental processes revealed by reaction-time experiments. Am. Sci. 1969; 57:421–57. [PubMed: 5360276]

Tillisch K, Mayer EA, Labus JS. Quantitative meta-analysis identifies brain regions activated during rectal distension in irritable bowel syndrome. Gastroenterology. 2011; 140:91–100. [PubMed: 20696168]

Turkeltaub PE, Eden GF, Jones KM, Ze ro TA. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. Neuroimage. 2002; 16:765–80. [PubMed: 12169260]

Turner JA, Laird AR. The cognitive paradigm ontology: design and application. Neuroinformatics. 2012; 10:57–66. [PubMed: 21643732]

Varoquaux G, Craddock RC. Learning and comparing functional connectomes across subjects. Neuroimage. 2013; 80:405–15. [PubMed: 23583357]

Wager TD, Atlas LY, Lindquist MA, Roy M, Woo CW, Kross E. An fmri-based neurologic signature of physical pain. N. Engl. J. Med. 2013; 368:1388–97. [PubMed: 23574118]

Wager TD, Lindquist M, Kaplan L. Meta-analysis of functional neuroimaging data: current and future directions. Soc. Cogn. Affect. Neurosci. 2007; 2:150–8. [PubMed: 18985131]

Wager TD, Rilling JK, Smith EE, Sokolik A, Casey KL, et al. Placebo-induced changes in fmri in the anticipation and experience of pain. Science. 2004; 303:1162–7. [PubMed: 14976306]

Wiech K, Lin Cs, Brodersen KH, Bingel U, Ploner M, Tracey I. Anterior insula integrates information about salience into perceptual decisions about pain. J. Neurosci. 2010; 30:16324–31. [PubMed: 21123578]

Wilhelm O, Hildebrandt A, Oberauer K. What is working memory capacity, and how can we measure it? Front. Psychol. 2013; 4:433. [PubMed: 23898309]

Wise RJ, Greene J, Büchel C, Scott SK. Brain regions involved in articulation. Lancet. 1999; 353:1057–61. [PubMed: 10199354]

Yarkoni T, Barch DM, Gray JR, Conturo TE, Braver TS. Bold correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fmri analysis. PLoS One. 2009; 4:e4257. [PubMed: 19165335]

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods. 2011

Yarkoni T, Poldrack RA, Van Essen DC, Wager TD. Cognitive neuroscience 2.0: building a cumulative science of human brain function. Trends Cogn. Sci. 2010; 14:489–96. [PubMed: 20884276]

Yeo BTT, Krienen FM, Eickhoff SB, Yaakub SN, Fox PT, et al. Functional specialization and flexibility in human association cortex. Cereb. Cortex. 2014
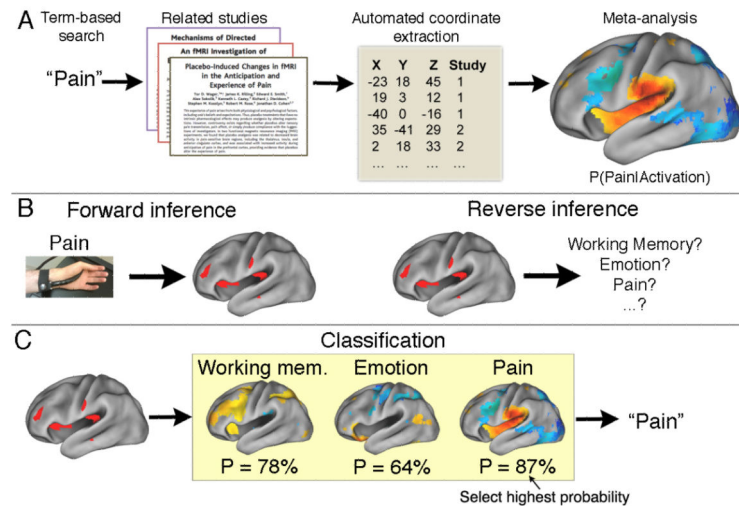
Figure 1.
An overview of the Neurosynth project (reprinted from (Yarkoni et al. 2011)). A) Neurosynth involves the automated extraction of activation coordinates from published papers, which can be used to perform a meta-analysis to identify regions associated with the presence of specific terms in publications. B) Forward inference involves estimation of the likelihood of activation given presence of a term, whereas reverse inference involves estimation of the likelihood of presence of a term given activation in each voxel. C) Neurosynth can be used to predict which of a set of terms is most likely to be present in a paper given a particular activation pattern.
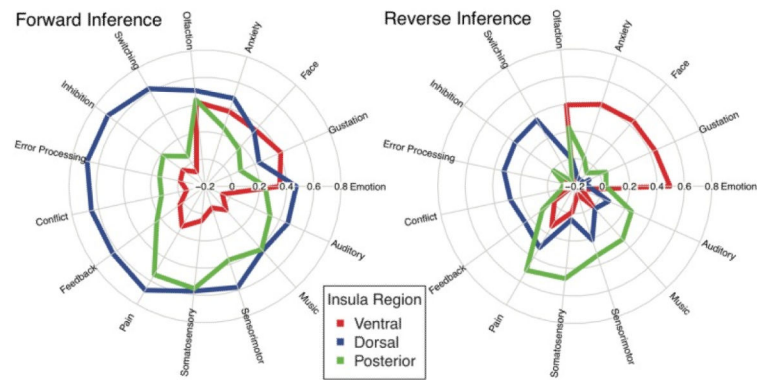
Figure 2.
Results of large-scale forward-inference and reverse-inference analyses of insula function in the Neurosynth database (reproduced from Chang et al. (Chang et al. 2013)). Whereas the dorsal anterior insula (blue) is consistently activated by a broad range of tasks in the forward inference analysis, the reverse inference analysis reveals it to be preferentially associated with higher cognitive functions.
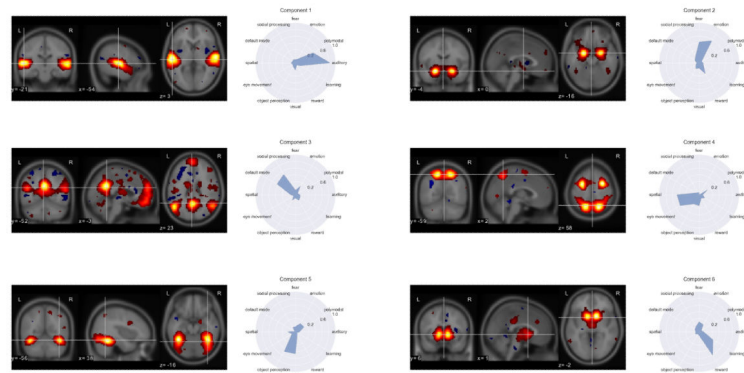
Figure 3.
Multivariate meta-analysis identifies mappings between neural activation and psychological function. Probabilistic independent components analysis (ICA) was performed on 9,721 images from the Neurosynth database, using MELODIC (Beckmann & Smith 2004). The slice maps show voxels that were significantly associated with each of the top six components (red-orange: positive association, blue:negative association). Polar plots show the relative association between each component and selected latent topics from the database (cf. (Poldrack et al. 2012)), demonstrating that most ICA components showed relatively selective associations with cognitive topics.
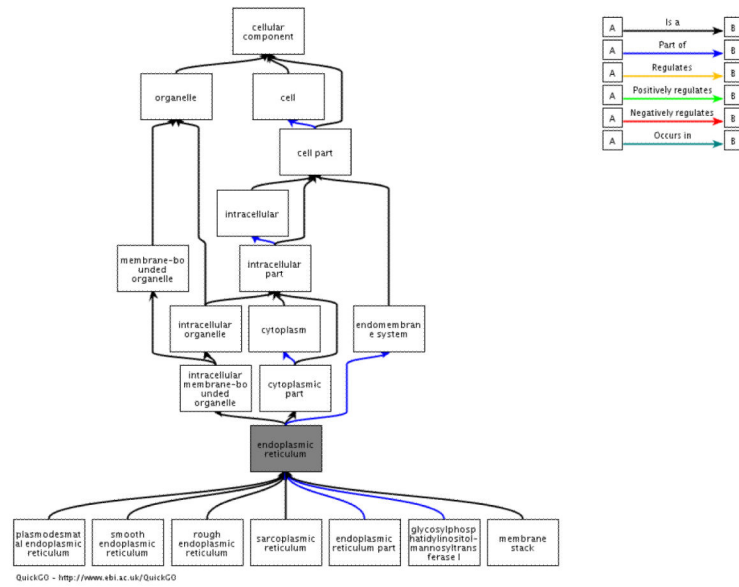
Figure 4.
An example of entities and relations within the Gene Ontology. This chart (generated using the QuickGO browser: http://www.ebi.ac.uk/QuickGO) shows the relations of the term "endoplasmic reticulum"; different possible relations between entities are denoted using different line styles.
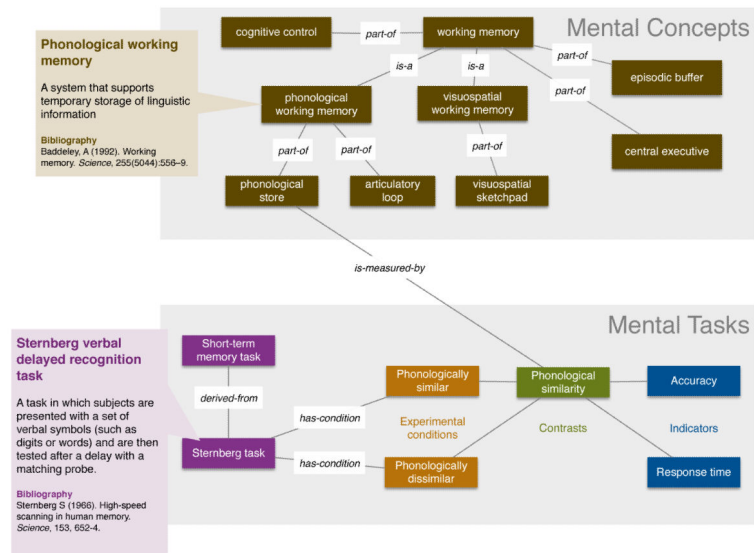
Figure 5.
A schematic example of the Cognitive Atlas representation of Baddeley's working memory theory. The top panel shows a representation of the concepts underlying the theory, whereas the bottom panel shows an example of a task measuring one of these concepts.