colleagues (2013) typically use a coarse-grained categorization scheme, distinguishing about 20 processes such as vision, attention, phonology, semantics, learning, or working memory. This scheme allows them to represent quantitatively how often, according to a given literature, a given area is activated when one of these 20 processes is recruited by an experimental task, for example how often articles studying working memory report activation in the dorsal anterior insula. The pattern of recruitment of a given area, given a particular set of fMRI articles and a categorization scheme, is its functional fingerprint. Although, unsurprisingly, areas tend to be activated by many processes, their functional fingerprints vary. Importantly, a functional fingerprint is a mere *summary* of a data set: It does not explain why the area is activated the way it is.

Following Poldrack et al. (2009), Anderson (2014, sects. 4.3 and 4.4) proposes to use dimension reduction techniques (factor analysis, MDS, PCA, etc.) to identify a few dimensions explaining why an area has its functional fingerprint. Instead of merely summarizing the involvement of a given area in a set of tasks, as functional fingerprints do, neural personalities *explain* this involvement: They allow cognitive neuroscientists to claim that *because* an area has a given neural personality (its score is *i* on NRP factor 1, *j* on NRP factor 2, etc.), it is involved more in some tasks than in others.

However, dimension reduction techniques are ill suited for discovering new cognitive constructs (Glymour 2001; Gould 1996). These statistical techniques project high-dimensional spaces onto spaces with fewer dimensions. On their own, the resulting dimensions cannot be interpreted realistically; they merely provide convenient ways of summarizing high-dimensional data. Three main arguments support this deflationary understanding of dimension reduction techniques. First, the outcome of these techniques is *underdetermined*. A given set of vectors in a high-dimensional space can be projected onto different spaces with different dimensions. To highlight merely three issues, there are many nonequivalent dimension reduction techniques, the number of dimensions is typically arbitrarily chosen, and these dimensions can be oriented in different manners. None of the possible spaces should be interpreted realistically because it would be arbitrary to treat one of them as real to the detriment of the others. Second, just like causally-based correlations, *accidental* correlations can be projected onto a lower-dimensional space, resulting in meaningless dimensions (e.g., Gould 1996, p. 280). Hence, that a high-dimensional space can be projected onto a lower-dimensional space does not justify interpreting the resulting dimensions realistically. Finally, the capacity of dimension reduction techniques such as factor analysis to identify causes has not been validated (Glymour 2001, Ch. 14). These three arguments bear on Anderson's project, exactly as they bear on IQ and personality research: On their own, dimension reduction techniques do not justify interpreting the dimensions of neural personalities realistically. Forgetting their limitations is committing the error of *reification* – namely, presuming that the abstract mathematical entities uncovered by dimension reduction analyses correspond to real psychological entities.

Naturally, the products of dimension reduction techniques can sometimes be interpreted realistically instead of as mere instruments for summarizing high-dimensional data. To do so scientists need to bring their broader empirical knowledge to bear on the interpretation of the dimensions of the lower-dimensional space. In the present context, this means that a purely bottom-up approach to cognitive ontology revision is unlikely to succeed: Some other information beyond the activation of brain areas across a range of tasks and their dimension reduction is needed to interpret the resulting dimensions realistically. Perhaps it is also worth noting that establishing the predictive validity of neural personalities does not justify understanding them realistically.

Anderson's approach to cognitive ontology revision is not the only one to fall prey to this *reification objection*; in fact, we speculate that in general purely bottom-up cognitive ontology revisions

commit the error of reification (e.g., Poldrack et al. 2009). Such approaches must reduce the very high-dimensional space defined by the number of voxels considered in order to identify cognitive constructs defined solely by brain activation patterns. Doing so probably requires using techniques whose product cannot be interpreted realistically. In our opinion, the reification objection reveals a fundamental shortcoming of bottom-up cognitive ontology revision.

Reason for optimism: How a shifting focus on neural population codes is moving cognitive neuroscience beyond phrenology

Carolyn Parkinson[a] and Thalia Wheatley[b]

[a]Department of Psychology, University of California, Los Angeles (UCLA), Los Angeles, CA 90095; [b]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755.
cparkinson@ucla.edu   thalia.p.wheatley@dartmouth.edu
http://csnlab.org
http://www.wheatlab.com

Abstract: Multivariate pattern analysis can address many of the challenges for cognitive neuroscience highlighted in *After Phrenology* (Anderson 2014) by illuminating the information content of brain regions and by providing insight into whether functional overlap reflects the recruitment of common or distinct computational mechanisms. Further, failing to consider submaximal but reliable population responses can lead to an overly modular account of brain function.

There is much to like in Michael Anderson's new book, *After Phrenology* (2014). First and foremost, he synthesizes a large and diverse literature to support a sophisticated and evolutionarily grounded approach to brain science. He describes the brain as a system in constant flux, transiently coalescing competing and cooperating neural assemblies in the service of action. The brain, he reasons, evolved to engage successfully with the environment rather than to represent the world accurately. He argues convincingly that this difference is underappreciated by cognitive neuroscience, which instead tends to look for neural activity that encodes objective properties in the environment as if they are context- and perceiver-free.

Anderson's description of the brain as a dynamic information processing device that responds to structured signals from the environment to guide action dovetails with his and other theories of neural reuse (Anderson 2010; Chang et al. 2013; Dehaene & Cohen 2007; Marcus 2006; Parkinson & Wheatley 2015). The basic idea here is that the brain solves new problems by repurposing preexisting neural architecture that solved structurally similar problems. This evolutionary and action-focused account of the brain is interwoven with the "extended mind" thesis, that intelligence is subserved by iterative interaction with the environment. These ideas form a solid, well-articulated, and reasonable book.

Less convincingly, Anderson argues that cognitive neuroscience is held in thrall to a "massively modular" account of the brain. From Anderson's vantage point, the field's modus operandi is to test for and report the single unified function of each brain area, as if the brain is a "collection of organs" (p. 292). As he points out, a brain composed mostly of special-purpose modules would be inefficient, prohibitively large, and inconsistent with the bulk of currently available evidence. We agree with Anderson that massive modularity is untenable, but so too does all but a vanishing minority of neuroscientists. The near-consensus view among contemporary neuroscientists is that most cortical regions exhibit considerable functional heterogeneity. Hence, the pleas to jettison modularity feel dated.

Anderson correctly asserts that the interconnected complexity and flexibility of neural activity presents immense challenges for brain science. The categories that we use to reflect on our own cognition are unlikely to match the functional organization of the brain, and a given brain region likely serves multiple functional roles, which can differ depending on the current environmental context and the networks in which the brain region is currently participating. However, we are somewhat more optimistic that the field can make significant headway in uncovering the structure of operations underlying human thought and behavior. This optimism is fueled in large part by more than a decade of cognitive neuroscience research that has honed techniques for multivariate pattern analysis of functional neuroimaging data. *After Phrenology* briefly alludes to this approach (Kriegeskorte et al. 2008a), but we believe that it is of particular relevance to Anderson's proposed framework and to his suggested ways forward for the field, and hence, merits further emphasis.

We now know that the brain encodes many kinds of information (e.g., memories, tastes, reach trajectories) in neuronal population codes, rather than in the magnitude of single units' responses (Georgopoulos et al. 1988; Lin et al. 2006; Pouget et al. 2000; Simon et al. 2006). Fortunately, much of the information contained in neuronal population codes can be gleaned from the indirect and spatially coarse measurements of brain activity that are currently available to researchers studying human brain function (e.g., functional magnetic resonance imaging, fMRI; Kriegeskorte et al. 2008b).

However, much of the information present in neural population codes is lost when analyzing the magnitude of responses of a given functional unit in isolation. For example, one of the first studies that analyzed fMRI data in terms of population response vectors demonstrated that brain regions that are typically characterized in terms of the categories of stimuli eliciting maximal responses (e.g., the fusiform face area) also differentiate between stimulus categories that evoke submaximal responses, such as chairs, shoes, cats, and houses (Haxby et al. 2001). This information was carried in unique and consistent topographic arrangements of responses for each stimulus category, and such information had previously been missed by analyzing data in terms of response magnitude only. Hence, analyzing neural responses only in terms of overall response magnitude can lead to an exaggeratedly modular conceptualization of brain function, as it can underestimate the number of brain regions involved in processing a stimulus category, as well as the number of stimuli a given brain region is involved in processing. Therefore, the brain may be even less modular than the functional fingerprinting method that Anderson employs would suggest, given that this method appears to only take into account response magnitude.

Analyzing data in terms of population response vectors, rather than overall response magnitude, can also provide insight into how information is organized within brain regions. To this end, researchers have applied both supervised and unsupervised machine learning techniques when analyzing the distributed topographies of activity within brain regions in order to characterize the distinctions that those brain regions make about various classes of stimuli. For example, by characterizing the similarity structure of population response vectors to visual stimuli within brain regions of the ventral visual pathway, researchers have gained insight into the distinctions that brain areas within that pathway make about sets of stimuli, and hence, the operations that they may contribute to visual processing (Connolly et al. 2012). Studies using unsupervised learning techniques (e.g., clustering, multidimensional scaling) have the notable benefit of being wholly data-driven in the structure that is revealed. Hence, such approaches will be useful in following Anderson's call for a science focused on illuminating the psychological factors that best account for observed differences in neural activity, rather than on attempts to map psychological constructs onto the brain.

Considering the information content of multivoxel response patterns, rather than focusing solely on response magnitude,

also allows researchers to probe the functional significance of overlap in the brain regions that are recruited to process diverse domains of contents. Univariate, subtraction-based approaches to analyzing brain activity are relatively well suited to asking questions such as, "Is this brain area involved in both Task X and Task Y?" and "Is this brain area recruited 'more' for Task X than Task Y?" In contrast, analyzing neural population responses is an approach that is relatively better suited to uncovering *how* a brain region might contribute to various tasks. Functional overlap can reflect the recruitment of shared or distinct mechanisms, and these possibilities can sometimes be disentangled by studying neural population response vectors. For example, although univariate, magnitude-based analyses of fMRI data suggested that most voxels in a region of lateral occipital-temporal cortex responded to both static pictures of bodies and movement displays, subsequent analysis of the population responses in that region demonstrated that the response patterns to these two sets of stimuli, though spatially overlapping, were unrelated (Peelen & Downing 2007). Observations such as these are consistent with Anderson's suggestion that a brain region's computational role can differ depending on the circumstances surrounding its recruitment (e.g., the processing demands of a stimulus, the other brain regions with which it is interacting). Hence, analyzing population response vectors can help determine whether spatially overlapping functional activity for two different tasks reflects the recruitment of the same or dissociable underlying mechanisms.

The analysis of population response vectors can also reveal when functional overlap reflects the recruitment of a common underlying neural operation across multiple experimental contexts. For example, a machine learning classifier trained to distinguish between population response vectors within the superior parietal lobule (SPL) associated with leftward and rightward eye movements can correctly distinguish between population response vectors in the SPL corresponding to mental subtraction and mental arithmetic (Knops et al. 2009). This suggests that the SPL contributes shared operations when shifting spatial attention within the external environment (during saccades) and within internal mental representations (when performing arithmetic). This likely reflects a case of neural reuse through which functional architecture with a preexisting role in shifting attention in the external environment is exploited in order to perform analogous operations on mental representations, and hence, to perform the culturally learned skill of arithmetic.

We recently used classification and similarity structure-based analyses of neural population response vectors to test for a common neural encoding of egocentric distance in spatial, temporal, and social frames of reference (Parkinson et al. 2014). Cross-domain decoding analyses, as well as analyses of the similarity structure of population responses, revealed that areas such as the right anterior inferior parietal lobule, which has a long-established role in representing physical space, organize information not based on distance category (e.g., "Is this a social or temporal distance?"), but based on distance from the self (i.e., "Is this relatively close to or far away from me?"). These results are consistent with Anderson's suggestion that the brain is largely concerned with computing behaviorally relevant information, such as distance from oneself (and hence, relevance for action). Further, these results are consistent with the possibility that neural mechanisms for encoding information about physical space may have been reused in order to operate on increasingly abstract contents, such as temporal and social relationships (Parkinson & Wheatley 2013; 2015).

Understanding how the brain gives rise to a given facet of cognition or behavior requires not only identifying the neural substrates involved, but also uncovering the kinds of operations performed by those substrates. We are optimistic that advances in neuroimaging methods, such as the analysis of neural population response vectors, will be useful in moving toward a deeper and more veridical understanding of functional brain organization and of the flexible and adaptive computations that underlie

everyday thought and behavior. In so doing, cognitive neuroscience has the potential to elucidate the particular functional capacities that may be reused over the course of evolution and development to solve novel problems.

Multisensory integration substantiates distributed and overlapping neural networks

doi:10.1017/S0140525X15001612, e127

Achille Pasqualotto

Faculty of Arts and Social Sciences, Sabanci University, Tuzla 34956, Istanbul, Turkey.

achille@sabanciuniv.edu

Abstract: The hypothesis that highly overlapping networks underlie brain functions (neural reuse) is decisively supported by three decades of multisensory research. Multisensory areas process information from more than one sensory modality and therefore represent the best examples of neural reuse. Recent evidence of multisensory processing in primary visual cortices further indicates that neural reuse is a basic feature of the brain.

Cognitive and perceptive functions are supported by highly overlapping neural networks distributed throughout the brain, and this phenomenon can be referred to as "neural reuse" (Anderson 2010; 2014; Pessoa 2012). To use a metaphor, we might imagine the brain as a choir and neurons as the singers composing the choir; when the choir sings Song 1, some singers do not sing at all while others sing vigorously. "Active" singers represent the portion of the choir performing Song 1 (see Fig. 1a); in terms of the brain, those singers represent the neural network (parts of the brain) activated during a given cognitive or perceptive process (e.g., visual perception). When the choir sings Song 2, a slightly different, but highly overlapping portion of the singers might be active (see Fig. 1b). This exemplifies how different tasks (Song 1 and Song 2) recruit highly overlapping, but not identical, portions of the choir (see Fig. 1c); in the brain, a given cognitive or perceptive task (e.g., visual imagery) recruits a neural network highly overlapping, but not identical, to the one associated with visual perception (Ganis et al. 2004).

One of the hot topics for debate in experimental psychology and neuroscience is on the extent of *specialisation* or *distribution* of the brain functions (Anderson 2010; Driver & Noesselt 2008; Liang et al. 2013). Empirical evidence seems to favour the idea that the brain operates in a distributed, rather than specific, manner. For example, brain areas that were considered to be specialised for specific tasks (e.g., the fusiform face area for face recognition) have been found to be activated during performance of other tasks (e.g., recognition of cars; Gauthier et al. 2000). Research has also shown that cognitive and perceptive processes usually involve networks of brain areas rather than one specific area; for example, a very specific process such as face recognition recruits a network of brain areas including the occipital, temporal, and frontal lobes rather than one specific portion of the brain (Goldstein 2009). This *distribution* of brain functions is not only seen for the processing of faces, but also for other perceptual and cognitive functions (if not all; Van Dijk et al. 2010; Yeo et al. 2011) such as perception (Takahashi et al. 2013; Uesaki & Ashida 2015), attention (Posner & Rothbart 2007; Shulman et al. 1999), memory (Alain et al. 1998; Desgranges et al. 1998), language (Duffau 2008; Horwitz & Braun 2004), spatial cognition (Burgess et al. 2001; Vallar et al. 1999), and body representation (Filippetti et al. 2015; Longo et al. 2010).

This phenomenon is not limited to the macroscopic level (i.e., brain areas), but has also been observed on the microscopic level. Some neurons have been found to respond to several types of stimuli, rather than to a specific type of stimulus, with different response patterns (e.g., firing frequencies). The difference in response patterns is the feature that distinguishes the stimuli represented by the neurons (Decharms & Zador 2000; Gerstner et al. 1997). In a simplistic example, during the processing of Stimulus A, Neuron 1 fires at a high frequency, Neuron 2 at a low frequency, and Neuron 3 at a medium frequency. Whereas, during the processing of Stimulus B, Neuron 1 fires at a medium frequency, Neuron 2 at a high frequency, and Neuron 3 does not fire at all. To reuse the above-mentioned metaphor, the same singers might participate differently in performance of different songs.

The notion that brain functions are based on distributed and overlapping neural networks is convincingly supported by the findings that input from different sensory modalities activate distributed and overlapping networks of brain areas – namely, multisensory processing (Ricciardi et al. 2014; Stein & Stanford 2008; Stein et al. 1988). Multisensory areas are portions of the brain processing input from different sensory modalities. In the last decades, an increasing number of multisensory areas have been identified (Amad et al. 2014; Gallese et al. 1996; Gobbelé et al. 2003; Sereno & Huang 2006), suggesting that the brain is more engaged in multisensory processing than was initially believed. Classic multisensory (or *associative*) areas are activated by visual, auditory, and somatosensory input and consist of prefrontal (Fuster 1988; Öngür & Price 2000), posterior parietal (Andersen et al. 1985; Serino et al. 2011), and superior temporal (Beauchamp et al. 2008; Bruce et al. 1981) cortices. More recently, new multisensory areas have been identified, in the posterior frontal (Grafton et al. 1997), temporoparietal (Matsuhashi et al. 2004), and occipitotemporal (Beauchamp 2005) cortices.
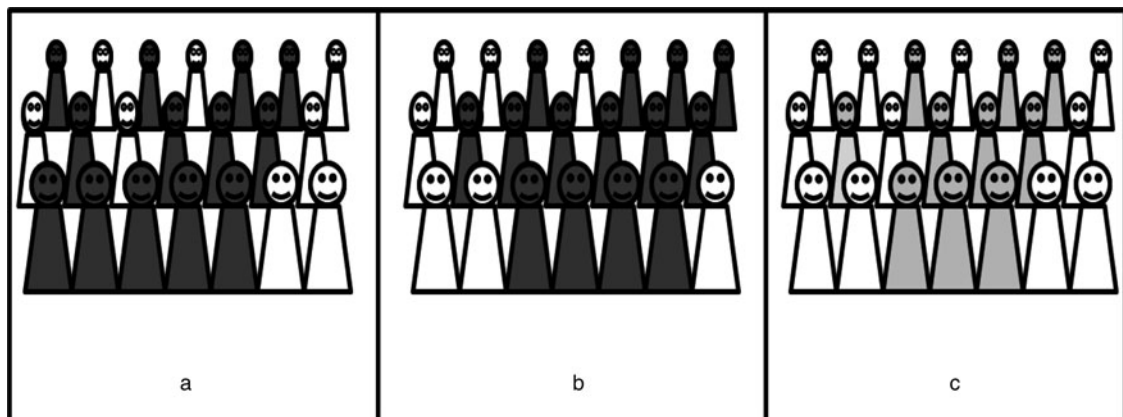

Figure 1 (Pasqualotto).   Singers of the choir singing (a) Song 1 (dark-grey); (b) Song 2 (dark-grey); (c) both Songs 1 and 2 (light-grey).