The Mirage of Big-Data Phrenology

Felipe De Brigard and Bryce S. Gessell

The goal of mapping psychological functions to brain structures has a venerable history. With the advent of neuroimaging techniques, this elusive goal regained vigour and became the main purpose of cognitive neuroscience. Unfortunately, as the field continues to develop, the ideal of finding one-to-one mappings from psychological functions to brain areas looks increasingly unrealistic. In the past few years, however, many cognitive neuroscientists have advocated for mining large sets of neuroimaging data in order to find this elusive one-to-one mapping. A recent strategy, proposed by Genon and colleagues, constitutes one of the most concrete proposals for discovering the mappings from brain regions to cognitive functions by using big-data repositories of neuroimaging results. In this article we offer several challenges for their proposal and argue that big-data approaches to finding one-to-one mappings between brain regions and cognitive functions suffer from significant difficulties of their own.

1. Introduction

The idea of mapping psychological functions to brain structures has a venerable history, dating back to Galen's ventricular doctrine (Green [2003]) and continuing to Gall's phrenology (Gall and Spurzheim [1810]). Although those theories are now in disrepute, the advent of neuroimaging techniques, such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), electro-encephalography (EEG), and magnetoencephalography (MEG), gives the prospect of finding one-to-one correlations between psychological functions and brain structures new vigour, and the project is the main goal of the young field of cognitive neuroscience (Posner and DiGirolamo [2000]).[1] Yet many doubt that cognitive neuroscience can give us such a psychological atlas, whereby the building blocks of mind get assigned to specific neural structures (Uttal [2001], [2011]). As cognitive neuroscience develops, the ideal of finding one-to-one mappings from psychological processes to brain areas (fig. 1a) looks more and more unrealistic, as the evidence increasingly suggests that mind-to-brain mappings are likely not one-to-one, nor even one-

[1] In this article, just as in the article we target (Genon et al. [2018]), the terms 'brain area', 'brain function', and 'brain structure' are used interchangeably.

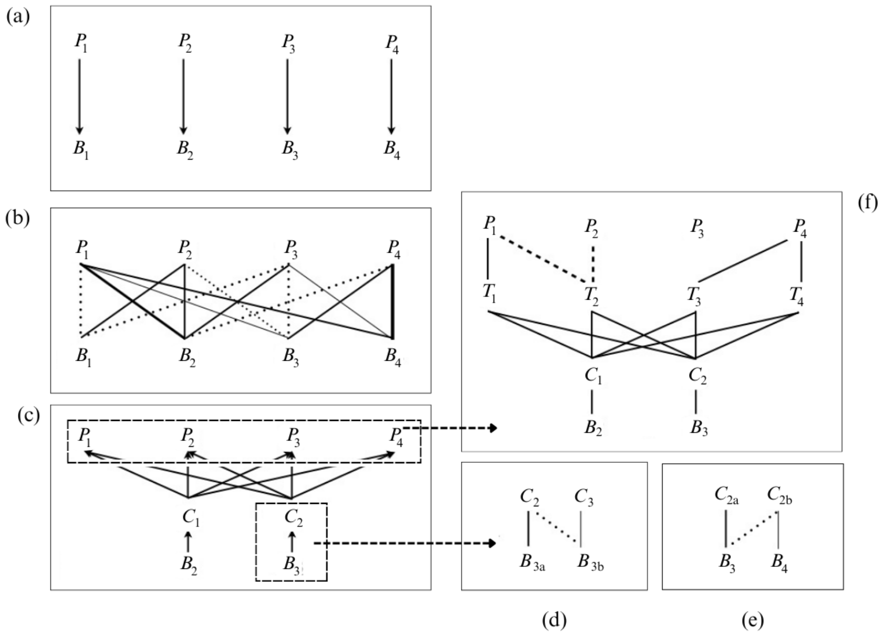

Figure 1. Models of the relationship between psychological functions (*P*) and brain regions (*B*). (a) Idealized one-to-one mapping of structure to function from a traditional top-down approach. (b) The many-to-many relations that have been uncovered through years of research in cognitive The differences in style and thickness of the lines represent different probabilities of between brain regions and psychological functions. (c) Idealized resolution with an intermediate computational layer involving operation functions, as advocated by Genon et al.'s ([2018]) approach. (d) Break down of a putative basic brain region into different sub-regions may one-to-one mapping from brain region to operation function, representing one problem for up approach. (e) Decomposition of a putative operation function into different computations remove the one-to-one mapping between brain region and operation function. representing problem for the bottom-up approach. (f) Mismatch between tasks and psychological function that emerge as a result of theoretical or methodological developments, representing one last problem for the bottom-up approach.

to-many (Barack [2019]; Viola [2017]; McCaffrey [2023]). Instead, it seems as though many diverse psychological functions are associated with the same brain regions, albeit perhaps with different degrees of probability, while at the same time many brain regions are engaged by the same psychological processes in different degrees (Price and Friston [2005]; Anderson [2014]; see also fig. 1b). Thus the goal of mapping psychological functions to brain structures appears to many people very unlikely, if not impossible (Fodor [1999]; Coltheart [2006]; Uttal [2001], [2011], [2012]).

Genon et al. ([2018]) argue that these difficulties stem from neuroscientists following a top-down approach: they start from accepted psychological categories (for example, object recognition, memory retrieval, sensorimotor integration) and then move down to the brain areas with which those categories are supposed to correlate. This approach, Genon et al. argue, is flawed and unlikely to yield the kinds of mappings cognitive neuroscience seeks. To solve the problem, the authors suggest a bottom-up approach, whereby researchers begin with the '*a priori* defined construct of the brain region' and work their way up to the 'unknowns', which are the psychological categories associated with that region. Their bottom-up approach begins by identifying a region of interest and then discovers the many cognitive functions, processes, and tasks—behavioural functions, in their terminology—associated with it. Although this strategy won't uncover one-to-one mappings, it should, they suggest, reveal the region's 'operation-function', or the computationally described function that 'grounds' all associated behavioural functions, but which 'remains latent and is not directly observed' (see fig. 1c).

Critically, given the limitations of traditional piecemeal experimental methods (for example, lesion deficits, single fMRI studies), Genon and colleagues favour using a big-data approach to identify a region's operation-function. By aggregating over thousands of studies—via, for instance, BrainMap (Fox and Lancaster [2002]), NeuroVault (Gorgolewski et al. [2015]), or Neurosynth (Yarkoni et al. [2011])—and by employing large-scale population samples (that is, HCP, UK, Biobank), Genon and colleagues claim we could identify every behavioural function associated with a particular region. Characterizing the operation-function would then involve finding a common computational role across all behavioural functions for each specific brain region.

We argue here that despite its promise, their proposed big-data bottom-up approach faces serious challenges. Section 2 begins by reconstructing Genon et al.'s arguments against the top-down approach. Section 3 reconstructs their proposed bottom-up approach. In section 4 we offer three challenges to their view and argue that their proposal doesn't fare better than the top-down approach they criticize. Section 5 gives an assessment whereby neither an exclusively top-down nor a bottom-up approach is favoured, but also a diligent piecemeal approach that requires, in fact, the clarification of basic conceptual confusions.

We must note, though, that Genon et al.'s proposal is not the only one that suggests a big-data bottom-up approach to overcome the difficulties with the top-down approach. In the past two decades other proposals have advocated for big-data bottom-up approaches to reforming cognitive ontologies, starting with Price and Friston's ([2005]) advice for 'good' ontologies, followed shortly by machine learning-based efforts like the Cognitive Atlas Project in cognitive neuroscience (Poldrack et al. [2009], [2010]), the Consortium for Neuropsychiatric Phenomics in neuropsychiatry (Bilder et al. [2009]), as well as more recent ones based on neural network analyses (for example, Yeo et al. [2015]; for reviews, see Anderson [2015]; Poldrack and Yarkoni [2016]). The hope is that as we learn more about the brain, we will have a way to clarify traditional concerns about cognitive ontologies—which, until recently, were debated entirely *a priori* or, at best, with mere behavioural data (Haueis [2014]; Anderson [2015]). We focus on Genon et al. here because their view constitutes one of the most recent and careful articulations of how the bottom-up approach, boosted by big data, might look in practice; but our arguments likely apply to other proposals as well.

2. The Top-Down Approach

To reconstruct Genon et al.'s argument against the top-down approach, we must familiarize ourselves with some of their terminology. First is the notion of behavioural function, which they employ as an umbrella term for all sorts of cognitive processes and operations identified by the many sciences studying the mind and behaviour, like 'episodic memory', 'motor preparation', 'visual attention', 'perspective taking', and 'emotion regulation' (Genon et al. [2018], p. 351).[2] By contrast, they use the term 'operation-function' to refer to 'a computational operation performed by a given region, which contributes to the observed behavioral output' ([2018], p. 352). Accordingly, every brain region is associated with many behavioural functions via a single latent operation-function—a core computational operation that unifies or 'grounds' the many behavioural functions but which, unlike them, 'is not directly observed. In other words, our current knowledge of the functional specialization of a given brain region can be conceptualized as a polyhedron with its many sides (that is, many behavioural functions), the sum of which can only be appreciated by

[2] An anonymous reviewer pointed out that the term 'behavioural function', as employed by Genon and colleagues, conflates both behavioural and cognitive constructs. We agree but will retain their terminology in this article.

investigation from many different perspectives, but whose core center remains intangible' ([2018], p. 352). As an illustration, the authors invite us to consider the hippocampus. This structure has been associated with many behavioural functions, including episodic and relational memory, recollection, encoding, retention, consolidation, novelty detection, pattern separation, and binding. Outside of memory, though, the hippocampus has been associated with behavioural functions as varied as spatial navigation, scene construction, prospection, episodic counterfactual thinking, and allocentric representation (Morris [2007]; Knierim [2015]). All these behavioural functions constitute the hippocampus's functional polyhedron; to ground all these seemingly disparate behavioural functions, there arguably exists a single operation-function that, unfortunately, is not directly observable. Now, according to Genon et al. ([2018]), researchers usually go about creating functional polyhedra by employing a top-down approach to mapping behavioural functions onto brain regions. But they argue that all known varieties of this approach are problematic. Here we reconstruct their criticisms and add some additional points.

The first variety of top-down approach is the lesion-deficit approach, whereby researchers characterize the function of a brain region based on the nature of the deficit the patient presents as a result of a brain lesion. Famous neuropsychological cases, such as that of patient Tan or H.M., epitomize this approach. Tan's lesion was taken as evidence that the function of Broca's area—the bit of neural tissue between the pars opercularis and the pars triangularis in the human prefrontal cortex (but see below)—was language production. Likewise, H.M.'s lesion became a critical piece of evidence that a central function of the hippocampus was to encode new episodic memories (cf. De Brigard [2019]). Dozens of other cases have been and continue to be used to make claims about functions of lesioned areas.

Unfortunately, as Genon et al. remark, the lesion-deficit approach is problematic. For instance, lesion studies are only quasi-experimental, so causal inferences are generally unwarranted, not only because pre-lesion factors cannot typically be ruled out, but also because the lesions themselves almost never respect neuroanatomical boundaries (Mah et al. [2014]). There are other well-known limitations of the lesion-deficit approach that Genon et al. don't mention but that are worth including here, such as the fact that it is hard to generalize dissociations across patients with similar neuropsychological profiles (Caramazza [1986]) and even within a single patient, unless one adopts several questionable assumptions (Shallice [1988]). Moreover, neuropsychological profiles are often incomplete and inaccurate, mainly because they are only as good as the tests characterizing them, and such tests often have shortcomings. Consider H.M., who allegedly had intact working memory even though this conclusion was drawn from a single digit-span test, which isn't an accurate measure of the entire construct of working memory (De Brigard [2019]; Boone and De Brigard [forthcoming]).

The second way to create functional polyhedra is the stimulation approach, which tries to more experimentally impair a brain region in order to better control the causal path inducing a deficit in a behavioural function. Stimulation approaches could be either invasive, such as deep brain stimulation and epidural motor cortex stimulation, or non-invasive, such as transcranial magnetic stimulation (TMS), transcranial electric stimulation (tES), and transcranial direct current stimulation (tDCS), each of which involves stimulation from outside the individual's cranium.

Unfortunately, these methods have shortcomings too. Invasive stimulation is practically difficult; DBS requires a complex surgical operation and thus typically involves non-neurotypical individuals, which not only raises concerns about pre-lesion conditions, but also makes it difficult to find suitable control groups. Epidural MCS involves less disruptive surgery but, as its name indicates, the method is topographically limited to a handful of cortical structures. Non-invasive brain stimulation techniques don't fare much better. TMS stimulates to a depth of around 3 cm, unsuitable for many sub-cortical regions. There are also concerns about concurrent sensory and motor side effects due to non-focal changes in magnetic fields. Finally, there are still questions about the how non-invasive transcranial stimulation affects brain regions at the micro-level, which complicates the chances clarifying how the stimulated brain region performs its putative function.[3]

[3] Genon et al. ([2008]) claim that the most serious concern with the stimulation approach is that, unlike the lesion-deficit approach, brain stimulation lacks ecological validity. But it isn't clear in what sense the lesion-deficit approach is more ecologically valid. One thing they may mean is that while lesions are naturally created, experimental brain stimulation isn't. But this would be a strange thing to say, not only because many brain lesions are manufactured—not only surgically, as in the case of H.M., but also accidentally, as in the case of traumatic brain injuries or accidental anoxic events—but also because lesions are almost never localized or limited by precise neuroanatomical boundaries. Or perhaps what they mean to say is that while we can see neuropsychological patients in a variety of mundane situations, brain stimulation is usually confined to awkward settings like surgical rooms or experimental labs. But again, a precise characterization of the behavioural effects of the lesion—whether permanent, as in the lesion approach, or transient, as in the brain stimulation approach—is task-dependent, and there is no more reason to believe that the tasks employed during TMS are less ecologically valid than those employed with neuropsychological populations. In fact, they are often the same! Thus, while we agree that there are difficulties with the stimulation approach, we disagree with Genon et al. about the main reasons.

The third top-down strategy is the activation approach, which makes use of recent neuroimaging techniques to identify neural activity associated with a particular task. These methods offer several advantages. For instance, fMRI, which relies on magnetic fields to track changes in deoxygenated haemoglobin, which in turn track metabolic increases in neuronal activity, enables researchers to record observations with a spatial resolution of a few millimetres (Huettel et al. [2014]). Similarly, PET—or fluorodeoxyglucose positron emission tomography (FDG-PET)—was first introduced to track glucose, another critical metabolic resource, thanks to injected isotopes that upon decay release positrons, which in turn collide with electrons emitting gamma rays for the scanner to detect (Raichle [1983]), although nowadays neuroimagers use other radioactive tracers for tracking other chemicals, primarily oxygen. By contrast, EEG and MEG do not offer comparable spatial resolution, but their temporal resolution is better than that of fMRI or PET, as they enable millisecond-sensitive recording through the scalp. Indeed, in recent years, researchers have been able to simultaneously record fMRI and EEG with some success (Huster [2012]). Given how minimally invasive these techniques are, and how much easier it is to control the independent variables in experimental designs, it is not surprising that activation approaches are so popular.

Unfortunately, like other top-down methods, they also have many limitations. Genon et al. ([2018]) mention a few, such as the problem of pure insertion, which affects many contrast-based studies where a particular cognitive process is allegedly ‘isolated’ by subtracting the other processes in common between two conditions (Friston et al. [1996]). They also note that many neuroimaging experiments have a hard time isolating task-specific effects, which would challenge the studies’ internal validity. To be fair, these problems are well known among neuroimagers—in fact, the problem of pure insertion arose even before neuroimaging (Sternberg [1969])—and many analytic methods, including multivariate and machine-learning analyses, have been put forth to overcome some limitations. Nevertheless, and in agreement with Genon and colleagues, it is no secret that neuroimaging techniques, despite such novel analytic methods, are still plagued with technical and conceptual difficulties (Gessell et al. [2021]; Boone and De Brigard [forthcoming]).

Finally, there is the structure–behaviour correlation approach, whereby biological characteristics of the brain—morphological, physiological, or volumetric—are correlated with behavioural measures across subjects or groups. Current technological developments have also made it possible to associate these brain features with other biological markers, such as genes and hormones, in many human populations. Moreover, recent developments in computational and analytic techniques have improved the predictability of these biological indicators on various behavioural measures (Kanai and Rees [2011]). Nevertheless, as Genon et al. ([2018]) discuss, structure–behaviour correlations are not immune to concerns about degeneracy: structurally different neural systems can yield indistinguishable behavioural performance (Price and Friston [2002]; De Brigard [2017]). Additionally, many neurological features are influenced by unknown factors that likely have little to do with the predicted variable, but since they cannot be regressed out, they could artificially inflate certain correlations (Westfall and Yarkoni [2016]).

Taken together, the challenges reviewed in this section suggest that extant varieties of top-down approaches to mapping behavioural functions to brain regions suffer from serious limitations. While a combination of techniques and our reliance on convergent evidence can help assuage some concerns, it is unlikely that every shortcoming can be eliminated. More importantly, the picture that the top-down approach offers today is very different from the one-to-one mapping it promised years ago (fig. 1a). Indeed, the neural picture we have gathered so far is one in which many psychological notions are associated, with varying degrees of probability, with many brain structures (fig. 1b).

What should we make of this? For Genon and colleagues, the root issue is a common assumption: that the starting point should be the '*a priori* defined construct of a mental operation', to then try to infer the brain region associated with it. For 'this *modus operandi*', they contend, 'has only a very limited capacity to answer the initial question: 'What does any part of the brain do'?' (Genon et al. [2018], p. 356). They suggest instead a bottom-up approach.

3. The Bottom-Up Approach

Overcoming the difficulties of the top-down approach, according to Genon et al. ([2018], p. 356), requires a complete change in perspective:

> Assessing the relative functional specialization of brain regions critically requires a change in viewpoint, where the *a priori* defined construct is the brain region, and the unknowns are the behavioral-functions associated with it. This implies screening a vast range of potential behavioral associations for a given brain region, and examining which of these are

> associated with the region of interest in an unbiased, statistically testable manner that accommodates the aforementioned complementarity of different approaches with respect to behavioral aspects.

The screening and statistical examinations of so many potential behaviour–brain associations would be difficult. Yet we can optimize the process, according to them, by using newly available ‘big data’ analytic tools. Genon et al.’s proposed bottom-up approach combines a different starting point—the *a priori* defined notion of a brain region—with the analytic advantages of big data to generate one-to-one mappings between brain regions and the operation-functions underlying the behavioural functions for the regions’ functional polyhedra.

The proposed bottom-up approach involves four steps. (1) First, identify *a priori* region of interest. Genon and colleagues offer two examples: the anterior insula and the hippocampus. (2) Next, use computational tools, like BrainMap and Neurosynth, to aggregate across all available studies where the activation maxima is reported as falling within the region of interest. (3) Afterwards, use statistical analyses with the entire dataset, while trying to account for both the region’s base rate of activation, as well as each behavioural condition associated with it.[4] In so doing, the thought goes, one should be able to identify particular behavioural operations consistently (that is, above some statistical threshold) associated with the region of interest. This step, which Genon et al. call ‘functional behavioral profiling’, is supposed to be tantamount to generating a data-driven functional polyhedron: its many ‘sides’, or the behavioural functions associated with it, are identified not by the limited amount of knowledge researchers possess but by the indefatigable thoroughness of a fancy big-data algorithm. (4) Finally, identify a generic functional role ‘that could account for all the more specific mental processes that have previously been discussed for this region’ (Genon et al. [2018], p. 357). This underlying functional role would correspond to the latent operation-function associated with that brain region.

Let us examine how this bottom-up approach might work with one of their examples: the anterior insula. Suppose we are interested in employing a bottom-up approach to identify the operation-

[4] One immediate problem is that researchers take enormous amounts of liberties in reporting coordinates and activation maps. Often the reporting is selective, and the coordinates of many results that survive thresholding go unreported. Given this, these base rates are likely very biased. We won’t return to this concern later in the article, but it is worth noting that reporting and base-rate biases are a major problem for any proposal seeking to employ big data in neuroimaging to generate behaviour–structure mappings.

function of the anterior insula. How would we go about doing so? Now that we have selected a region of interest (1), we then employ the most cutting-edge computational tool available to collate all extant published studies where the activation maxima fall within the anterior insula (2). Next, we query the tool for activation maps centred in the anterior insula, thereby outputting a list of several hundred studies with associated activation maps (3). Now we run some statistical analyses to eliminate studies where the association between the anterior insula activation and a behavioural function in that particular experiment does not survive some conservative threshold, minimizing the chances that the associations are capturing noise rather than signal. And though we do not currently have a tool that collates data from all available cognitive neuroscience studies, we do have a large-scale platform, Neurosynth, that can help us cull a good portion (14,371 as of 20 May 2023) of all published fMRI studies. Indeed, querying Neurosynth with an anatomical label of the anterior insula for all studies with reported activation in that area produces around 700 hits.

Now assume that the tool not only outputs the studies, but also gives us a curated list of the cognitive processes each study was set to identify. Just looking at the first few entries of a Neurosynth query, for instance, we get a range of behavioural functions, including ‘sustained attention’, ‘inhibition’, ‘executive function’, ‘regret’, ‘food motivation’, and ‘unreciprocated cooperation’, to name just a few. The thought is that after all these (big!) brain data are analysed, all the associated behavioural function terms curated, and the associations between them appropriately thresholded, we will then be able to

> […] identify a generic functional role, such as *task engagement maintenance*, that could account for all the more specific mental processes that have been previously discussed for the [anterior insula]. As illustrated in this example, the patterns of associations across a wide range of tasks can foster new hypotheses, approximating as much as possible the core role of the region (and thus its operation-function), beyond the behavioral ontology of the original studies or the database (Genon et al. [2018], p. 357; our emphasis).

Just to emphasize: Genon et al. are aware that we currently do not possess automated tools to aggregate over all extant cognitive neuroscientific studies, and that the best current datasets are limited to fMRI and PET studies. However, they do believe that, in time, their bottom-up approach will deliver accurate and systematic functional behavioural profiles for each brain region. From there, researchers would be able to determine the operation-function grounding all the behavioural functions comprising each individual brain region’s polyhedron (4). As a result, the daunting many-to-many picture (fig. 1b) threatening the viability of a cognitive ontology mapped cleanly to brain structures would be rendered tractable (fig. 1c) by the bottom-up approach.

4. Difficulties with the Bottom-Up Approach

While not unprecedented, Genon et al.'s ([2018]) proposal is one of the most serious, careful, and concrete attempts at showing how big-data approaches could reform structure–function mappings and cognitive ontologies. In this section, however, we raise three problems for their proposal, and argue that the problems threaten the viability not only of their project, but of any project using big brain data to map structures to functions or create cognitive ontologies. Specifically, we argue that big-data approaches face the problem of defining brain regions, the problem of model dependence, and the problem of the task-process barrier.

4.1. What's a brain region?

A fundamental assumption of Genon et al.'s proposal is that there are basic brain regions that can be defined *a priori*; the identification of one such region is, after all, the first step of the bottom-up approach. Unfortunately, the authors fail to define what a brain region is, how it can be identified *a priori*, and how to distinguish those brain regions that can be identified *a priori*—and for which it makes sense to work out a functional analysis according to the bottom-up approach—from those that cannot. They do offer two examples, however: the anterior insula (AI) and the hippocampus. According to the authors, these two structures, traditionally delimited neuroanatomically, appear to be clear instances of what they call '*a priori* basic brain regions' for which it makes sense to identify a single operation-function. But why should they be so?

To evaluate the claim that a structure like AI constitutes an *a priori* basic brain region in the sense assumed by the bottom-up approach, let us delve a bit into its neuroanatomy and neurophysiology. Macroscopically, the insula is a portion of the neocortex folded inside the lateral sulcus, covering between two and four percent of the total cortical surface, and enclosed dorsally by the frontoparietal operculum and ventrally by the temporal operculum (fig. 2a). The nomenclature 'anterior' versus 'posterior' insula comes from Brodmann's initial characterization in 1909 (fig. 2b). The two gyri dorsal to the sulcus circularis comprised the posterior insula, which was then characterized as granular, given its proportionally larger number of Meynert cells in layer 4. By contrast, the agranular AI comprised the three gyri rostral to the sulcus circularis.[5] Despite the fact that we keep employing this nomenclature today, only two years after Brodmann's initial cyto-arthitectonic parcellation, Vogt ([1911]) offered a different myelo-architectonic division between dorsal granular and ventral agranular sectors, roughly alongside the insula's sulci. Brockhaus's ([1940]) subsequent cyto- and myelo-architectonic parcellation identified an allocortical sector with two agranular regions: a mesocortical sector with eight dysgranular regions, and an isocortical region with sixteen granular regions. That's 26 cyto- and myelo-architectonically distinct regions within the insula—at least eight of which fall squarely into what Brodmann labelled AI (fig. 2c).

More recently, and employing more advanced staining and tracing methods, researchers have identified several cyto-architectonically distinct sub-regions in the macaque insula, with at least three of them in the dysgranular and seven in the agranular regions corresponding to the AI (Evrard et al. [2012]; Evrard et al. [2014]). Although that level of precision in the cyto-architectural organization of the human anterior insula is still lacking, our most current evidence suggests that it is probably as parcellated, if not more so, than that of the macaque (Bauernfeind et al. [2013]). Also, the AI is one of the very few cortical structures with atypical von Economo and fork neurons, whose selective degeneration has been recently associated with behavioural variants of frontotemporal dementia (Kim et al. [2011]). These neurons happen to be closely grouped in layer 5b of the anterior agranular insula, corresponding to approximately 3% and 1% of the total number of layer 5 neurons, respectively (Evrard [2018]; see also Evrard [2019]; Krockenberger et al. [unpublished]).

[5] There is typically a fourth accessory gyrus that varies in size and localization across individuals, which is already an important wrinkle for the bottom-up approach, though we will sidestep the issue here.

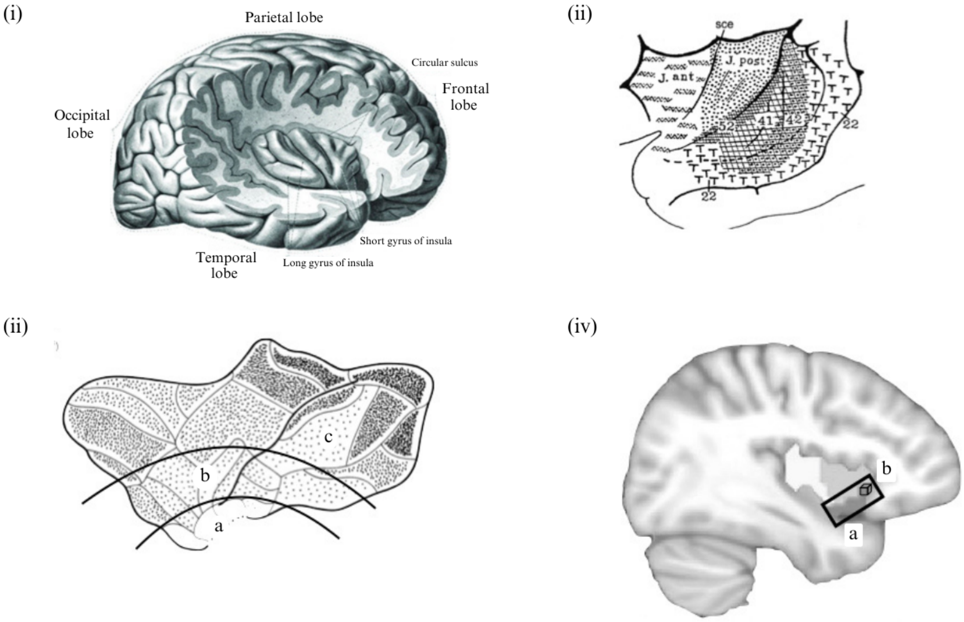

Figure 2. Neuroanatomy and neurophysiology of the human insula. (i) Neuroanatomical rendering of the insula operculum is removed. According to Brodmann's ([1909]) initial characterization (ii), the two gyri dorsal to the comprised the posterior insula, whereas the anterior insula laid rostral to the sulcus. (iii) Brockhouse's ([1940]) cyto-map of the human insula. The concentration of black dots indicates cortical granularity. The sector indicated by corresponds to the agranular portion of the anterior insula, while the sector under (b) roughly corresponds to the portion. The sector indicated with (c) comprises the granular portion, overlapping part of Brodmann's anterior parcellation. (iv) The region enclosed (a) covers the rostro-ventral agranular portion of the anterior insula, overlaid on functional parcellation map from a cluster analysis. It has a volume of approximately 2 cm^3, and is located below the granular zone. (b) Depicts—not to scale—an isotropic functional voxel. Depending on the resolution of the MRI could easily be 100 to 200 such voxels within the area demarcated by (a).

Now consider figure 2d, depicting the results of a cluster analysis of neuroimaging data identifying three sub-regions of the right insula (Deen et al. [2011]), and focus on the region enclosed by the rectangle, which very closely matches the ventral agranular portion of the AI, as identified by Brodmann (the same sector for which Brockhause ([1940]) found about eight cyto-architectonically distinct regions, and for which Evrard et al. ([2012]) found at least ten in the macaque). Volumetrically, this portion of the AI is approximately 2 cm^3 (Zhang et al. [2014]). Depending on the MRI scanner's resolution, between 100 and 200 isotropic voxels can fit within the enclosed region (fig. 2d). Now, according to our best estimations, a typical 3 mm functional isotropic voxel contains about 630,000 neurons (Herculano-Housel [2009])—although this number could be lower for agranular zones. Thus, even if we have a very high-resolution scanner, capable of giving us 1 mm functional isotropic voxels, and even if we assume that the proportion of neurons in the agranular portion of the anterior insula is half of that of its granular counterpart, there are still going to be over 100,000 neurons per voxel, and well over 100 voxels inside the enclosed region in figure 2d. And this does not even take into account glial cells, which are likely four times as many. That's a lot of brain cells.

This excursus into the intricate physiology of the insula seeks to motivate several pressing questions. First, what reason do we have to believe that the piece of cortical tissue known as AI, which we demarcate by gross neuroanatomy, is a good candidate for the sort of basic *a priori* brain region the bottom-up approach takes as a starting point? Moreover, given the aforementioned cyto-architectonic boundaries within the AI (Evrard [2014], [2018]), the fact that many of these sub-regions seem to project to different cortical targets (Krockenberger et al. [unpublished]), and the curious and yet not-well-understood fact that some of the most atypical neurons in the human brain happen to co-exist in a small portion of the AI, what reason do we have to believe that we can gloss over this complexity and accept that the AI is an ideal candidate for a basic *a priori* brain region—to which there corresponds a single operation-function? More dramatically—and this is a point we will discuss at length in the next section—what are the chances that the intricate biology housed inside the anterior insula manages to coalesce into a single computational operation that may be best described, according to Genon et al., as 'task engagement maintenance'? Albeit seemingly rhetorical, these questions should highlight a fundamental concern with the bottom-up approach: that it is not clear why the AI is a good candidate for a basic brain region, or what makes the selection of the AI as a starting point so obviously *a priori*. And what goes for the AI goes for any other allegedly basic *a priori* brain region, as we discuss later.

Before we get to the issue of what Genon et al. may mean by '*a priori*', it is worth discussing three additional complications with determining what a brain region is. The first one stems from what has been called the 'brain atlas concordance problem' (Bohland et al. [2009]; Ward [2022]). If you have ever worked with imaging data, you might have noticed that if a certain coordinate falls under a particular brain region when one parcellation protocol is used, it often falls under a different one when an alternative atlas is employed. There are several reasons why this occurs, not only having to do with the aforementioned difficulty of how to precisely delineate cyto-architectonic boundaries, but also with the fact that functional parcellation is probabilistic, and different available atlases employ different strategies to assign activation coordinates to particular brain regions. For instance, as Bohland et al. ([2009]) showed, one can have a relatively vast activation cluster that would get labelled as 'superior temporal gyrus' by the International Consortium for Brain Mapping (ICBM) anatomical template, but as 'medial temporal gyrus' by the Automated Anatomical Labeling (AAL) atlas. And while both the ICBM and the AAL atlases are widely used, they are not the only ones available, and other authors have reported similar concordance problems with alternative tools. Therefore, contrary to Genon et al.'s assumption that aggregating over hundreds of neuroimaging studies will give us an 'unbiased' way to determine associations between brain regions and behavioural functions, the fact is that whether a particular activation is reported as falling in one specific brain region is already biased by the parcellation protocol.[6]

It isn't unusual to try to solve concordance problems by visual inspection. Experts may look at where the peak activation is located and then, on the basis of their neuroanatomical knowledge, judge whether it falls in one or another region, thus overriding the automated labelling of their parcellation protocol. But this brings out a second problem: experts disagree. Recently, Tremblay and Dick ([2016]) asked 159 expert neuroanatomists from the Neurobiology of Language Society to give precise locations of two of the most important brain regions for research in the cognitive neuroscience of language: Broca's and Wernicke's areas. Surprisingly, their results revealed substantial disagreements about the extension and about which portions of cortical tissue should be included within each functionally defined area. Only 50% of respondents, for instance, stated that Broca's area includes the pars triangularis and the pars opercularis in the frontal cortex, with the other 50% selecting some other variation. Critically, though, within this second half of respondents, there were some drastic differences: 5% located Broca's area as being limited to the pars triangularis, 8% only included the dorsal portion of the pars triangularis, and 3% did not include the pars triangularis at all, confining it to the pars opercularis only. This means that if the peak

[6] The concordance problem occurs with pre-processed data, but the story could be even further complicated by the fact that different co-registration strategies in fMRI can yield differences in coordinate localization in three-dimensional space (Kashyap et al. [2018]), which in turn would yield differences in neuroanatomical labelling. Once again, this highlights the fact that all of these 'researcher degrees of freedom' inject a substantial amount of bias to what the bottom-up approach takes as the starting point of choosing a basic, *a priori* brain region.

maxima of a language task, say, falls in the rostral portion of the pars triangularis, more than 10% of expert neuroanatomists working in the neurobiology of language (assuming the sample is representative) are going to drastically disagree with their colleagues as to whether such activation falls in Broca's area. The verdict of experts may override concordance problems produced by different parcellation protocols, but their verdicts, too, could be biased.

Finally, a third difficulty is that the few examples of basic brain regions offered by Genon and colleagues are all topographically unified. However, it is reasonable to wonder, following recent proposals by Fox and Friston ([2012]) and Anderson et al. ([2013), whether we should assume, *a priori*, that the basic unit of brain structure must be topographically unified, as opposed to disaggregated bits of neural tissue working together. Just like Genon et al., these authors agree that the one-to-one mapping (fig. 1) many cognitive neuroscientists expected to bring about is unlikely, and that the reality looks more like a many-to-many mapping (fig. 1b). And just like Genon et al., they also advocate for the use of big-data and meta-analytic approaches to build a data-driven ontology. However, unlike Genon et al.'s bottom-up approach, these alternative proposals jettison the idea of starting with a neuroanatomically individuated brain region and instead suggest using integrative techniques—for example, functional connectivity, tractography, dynamic causal modelling, network analysis—to let the neural data determine what are the basic brain structures we should associate with cognitive functions. *Contra* Genon et al.'s bottom-up approach, these integrative proposals argue that the basic brain structures are likely not topographically unified but, rather, some sort of spatially and/or temporally distributed brain network. If these integrative approaches are correct and it turns out that topographically unified brain regions are not the basic structures of analysis, then the bottom-up approach will lead to situations where we would attribute the wrong operation-function to a brain region, either because it contributes to more than one operation-function (fig. 1d) or because it is part of a larger, basic brain region that contributes to a single operation-function (fig. 1f). Either way, the ideal one-to-one mapping from brain area to operation-function, once again, breaks down.

Perhaps some of these concerns hang on what exactly Genon et al. [2018]) mean by '*a priori*'. Although they never clarify the expression, they likely use it the way cognitive neuroscientists do when describing certain statistical analyses, that is, as synonymous with 'hypothesis' or 'theory-free'. But as we argued in this section, there is no such a thing as a hypothesis or theory-free way to determine what a brain region is. Even when it comes to cyto- or myelo-architectonic parcellation, the mapping will always be—as comparative neuroanatomist Evrard ([2014]) puts it—'biased by the subjective judgment of the observer and complicated by the inter-individual variability'. A neuroscientist's theoretical commitments penetrate their observations at many levels, including the fundamental one of choosing ideal candidates for determining which brain structures should be associated with operation-functions and which should not. Why did Genon et al. pick AI and hippocampus as good candidates for basic brain regions? Precisely because of their theoretical commitments with certain ways of neuroanatomically parcellating the brain, the resolution of current neuroimaging instruments, and years of top-down research in which those regions are taken to be good candidates for basic units of brain function. There is nothing '*a priori*' about the selection of these regions as starting points for a bottom-up approach, and there is nothing unbiased in the way their approach promises to deliver data-driven functional polyhedra.[7]

4.2. The model-dependency of operation-functions

In the previous section we challenged Genon and colleagues' assumption that there is a straightforward way to identify, *a priori*, the basic brain areas demanded by their proposed bottom-up approach—that is, our discussion concerned the level $B_1, \ldots, B_n$ in figure 1. In this section we challenge the next level, namely, that of identifying the computation that best characterizes the

[7] An anonymous reviewer suggested a possible amendment to Genon's et al. ([2018]) proposal to the effect that one could use alternative methods for selecting basic brain areas, and suggested, as an example, the cortical area parcellation approach taken by Gordon et al. ([2016]). Although this proposal is intriguing, there are two reasons why it may still not solve the issues discussed in this section. First, these kinds of approaches for cortical parcellation are full of theory-laden decisions. In Gordon et al. ([2016]) choices as to which levels of signal-to-noise ratio in the BOLD signal were appropriate, for instance, or which was the minimum size for a parcel (in their case, fifteen cortical vertices), were deliberate and guided by the theoretical decisions of the research team. At least currently, we don't think there is a fully theory-free way of conducting brain parcellations with imaging data. Second, even if there was, many of the issues having to do with resolution mentioned earlier in this section would still apply, since these parcellations are still voxel and BOLD-signal dependent. The 'real' basic units of the brain may be smaller than the resolution afforded by fMRI.

operation-function of an alleged basic brain region—that is, the level $C_1,\ldots,C_n$ in figure 1. Specifically, we argue that Genon and colleagues' proposed strategy to computationally characterize the operation-function of a brain region rests on three assumptions that are likely false.[8]

To that end, consider the second brain region they discuss: the hippocampus. Incidentally, it is curious that they picked this medial-temporal structure, which has been studied for over 400 years, because there is still disagreement as to what the label 'hippocampus' denotes. Some researchers think the term covers six regions: dentate gyrus (DG), cornu amonis (CA, which itself comprises three sub-regions CA1, CA2, and CA3), subiculum, presubiculum, parasubiculum, and entorhinal cortex (EC). But others disagree and think that the term should only cover allocortical structures—that is, evolutionarily older regions of the cortex that comprise fewer than six cortical layers—which would exclude the last three. And still others think that the hippocampus proper should only include CA1, CA2, and CA3, due to the nature of their enclosed axonal projections and granularity (Amaral and Lavenex [2007]). So, depending on the researchers' preferred terminology, we are talking about a structure that can vary from approximately 2 to approximately 5.2 cm, in humans, along its transversal axis (McHugh et al. [2007])—which, from the point of view of the bottom-up approach, should already give us pause. But let's set this issue aside and assume that 'hippocampus' refers to the formation including the six regions just listed.

How would we then go about determining the operation-function of the hippocampus? According to the bottom-up approach, and as detailed in section 3, the next step would be to employ a comprehensive algorithm to search through big-data repositories and cull all the results associated with hippocampal activation in order to generate its behavioural profiling. Again, as with the insula, if one tries to do so with Neurosynth today (20 May 2023), the algorithm outputs 1059 studies, which are certainly only a fraction of all imaging papers reporting activations in the hippocampus. Yet, this subset allows us to imagine what the behavioural profiling of the hippocampus would look like, were it to be generated by a much more comprehensive programme: a long list of articles reporting hippocampal activation across a variety of tasks and behaviours. Just going over the first ten hits, corresponding to the studies with the highest loadings, we already get a list of

[8] Of note, the view according to which a single brain area is associated with a single computational function is dubbed 'computational absolutism' by Burnston ([2016]). His criticisms of such a view are spot on, and entirely consistent with our considerations here. For an opposing view, see (Shine et al. [2016]).

quite diverse behavioural functions: ‘semantic memory’, ‘topographic memory’, ‘sequential reasoning’, ‘conditioning’, ‘extinction’, ‘episodic memory’, ‘encoding’, and ‘old/new judgments’. Genon and colleagues admit that big-data repositories and their culling algorithms are still in their infancy, yet they think that as they develop and include more and more studies, the emerging patterns of association will allow us to ‘generate hypotheses that will *approximate* as much as possible the core role’ of the hippocampus, from which in turn we will be able to ‘*infer* its operation [that is, computational] function’ (Genon et al. [2018], p. 357; our emphasis).

Unfortunately, the authors never explain what they mean by ‘approximate’ or what exactly is the nature of the inferential process allowing us to identify the operation-function of a selected brain region. A charitable reading indicates that this ‘approximation’ occurs because as the amount of data included in the generation of the behavioural profiling of the hippocampus increases, the number of hypotheses that could possibly characterize the computation of its core operation-function should decrease. However, the evidence suggests exactly the opposite: that as the number of studies showing hippocampal involvement increases, so does the number of computational models proposed to explain the operation-function of the hippocampus. Forty years ago, when computational neuroscience was emerging as a distinct discipline, there were only a couple computational models of hippocampal function. There was Marr’s influential ‘simple memory model’ of the hippocampus, for instance, which was part of his larger computational theory of the archicortex (Marr [1971]); the main objective was to offer a computational account for the hippocampus-neocortex interaction during memory encoding and retrieval. There were also computational models (for example, Zipser [1985]) to explain navigation behaviour and spatial location, thanks to the discovery of hippocampal place cells (O’Keefe and Dostrovsky [1971]). But the last thirty years have seen an explosion of computational models for hippocampal function, in no small part because the hippocampus has been shown to be involved in all sorts of behaviours via all sorts of different tasks. Moreover, there are competing computational models, involving different assumptions and parameters, vying to offer better fits for the same behavioural and neural data. In fact, there are now so many models that it isn’t hard to find entire chapters and volumes dedicated to computational models of hippocampal function for a single family of behaviours (for example, Gluck [1996]; Burgess [2007]; Hasselmo et al. [2020]). Contrary to Genon et al.’s assumption, then, it seems that the more behavioural functions the hippocampus is associated with, the greater the number of hypotheses for what its operation-function may be.

This shouldn't be surprising, though, for the proliferation of computational accounts of hippocampal function has to do with the nature of computational modelling itself. In essence, a computational neural model consists in a formal or mathematical representation of the mechanisms responsible for the operations of a brain system as well as the way such mechanisms interact to bring about certain cognitive, behavioural, or physiological process (Moustafa [2017]). As a result, computational models include parameters and variables ranging over quantifiable measures, some of which are physiological (for example, BOLD signal, dopamine release, and so on) and some behavioural (for example, Hit rates, RTs, saccades, and so on). The relative success of a computational model for a brain structure or process is ascertained based on how well it 'fits the data', that is, how closely it aligns with observed data and how precisely it generalizes to unobserved data. The issue, though, is that the data with which the fit of a model is tested are task-dependent. A computational model of reinforcement learning in the hippocampus, for instance, is going to be evaluated by how well it fits the data produced by a reinforcement learning task. Likewise, hippocampal models of relational memory encoding or spatial navigation would be tested against data produced by a memory or spatial navigation task. So, it is not surprising that as evidence accumulates showing hippocampal engagement during several different tasks, there is also an increase in the number of models offered to account for such findings in computational terms.

It sometimes happens, of course, that the same model can be employed to fit data from more than one study or more than one task (for example, Krasich et al. [2023]); this is indeed the best way to show that a model has explanatory breadth. But again, such tasks are usually thought to measure the same or very related cognitive process or behaviours. What is rarer is to find a single computational model that can account for two or more disparate set of findings, produced by different tasks, each tapping at theoretically different processes or behaviours. A recent example of this phenomenon is the computational model of the hippocampus as a predictive map (Stachenfeld et al. [2017]), which was proposed to solve two computationally conflicting models: one of the hippocampus as a cognitive map (O'Keefe and Nadel [1978]) and one of the hippocampus as a reward predictor in reinforcement learning (Schultz et al. [1997]). More recently, Whittington et al. ([2020]) offered a computational model to unify accounts of the hippocampus as a cognitive map and computational models of the hippocampus in relational memory task; there are thus instances where a model offers convergence between two or more disparate computational accounts of a brain structure. But the fact remains that these tend to be more the exception than the rule. And, critically, such convergent computational accounts are neither obvious nor the result of any straightforward inference. Thus, contrary to what Genon et al. intimate, there is little reason to believe that a more comprehensive behavioural profiling of a brain region is going to automatically make it easier to computationally characterize its operation-function.

But let us assume, for the sake of argument, that Genon et al.'s ([2018]) intuition is correct, and that the more data is included in the behavioural profiling of a brain region, the fewer the hypotheses as to what its operation-function may be. However, even if this was so, the vagueness with which Genon et al. characterize how to infer the operation-function of a brain region is problematic in a way that is reminiscent of concerns associated with the related proposal by Price and Friston ([2005]). In their widely cited article, Price and Friston, too, worry about the seemingly undeniable fact that each brain region appears to be associated with multiple cognitive operations. As a result, they advocate for a new computational ontology in which each brain region is associated with a functional characterization couched at the right level of abstraction. They use, as an example, the fact that the posterior lateral fusiform (PLF) gyrus is associated with a plethora of cognitive processes—not unlike the case of the AI or the hippocampus in (Genon et al. [2018])—but suggest that if one can find a computational label at an adequate level of abstraction so that it 'explains all patterns of activation', one can generate a computational characterization broad enough so that it will be 'more useful than task-specific labels' (Price and Friston [2005], p. 268). This new computational ontology is perhaps just what the bottom-up approach advocated by Genon and colleagues needs.

The problem for Genon et al. is that the path to a new cognitive ontology couched in abstract computational terms is deeply problematic. As Klein ([2012]) persuasively argued, if the only way to accommodate all the tasks associated with a brain region is to offer a computational label sufficiently abstract to cover them all, then it may end up with a placeholder label that most likely won't be particularly helpful. Price and Friston ([2005]) suggest that 'sensorimotor integration' could be a good characterization of the computational role of PLF, not unlike the 'task engagement maintenance' operation-function that Genon et al. ([2018]) suggest for the AI. But both labels are so general that do little to guide our cognitive theorizing. As Klein ([2012], p. 955) puts it:

> Specific functional attributions, when available, provide relatively strong constraints on cognitive theories. The more we abstract away from those details, the less constrained our cognitive theorizing becomes. To put it more bluntly: suppose we see PLF activation and so know that there is some sensorimotor integration going on. It is hard to know what cognitive theory could possibly conflict with that: at that level of abstraction, any theory looks like it will be compatible with PLF activation.

And the problem, of course, is that this concern is entirely replicated in the bottom-up approach proposed by Genon and colleagues.[9]

Finally, it is worth mentioning a third concern with the way the bottom-up approach suggests we should characterize computational functions. As stated, the suggestion is that as the behavioural profiling of a brain region becomes more and more comprehensive, the unique computational function of the *a priori* basic brain region will become clearer. However, this assumes that there is a one-to-one correspondence between a particular brain area and the right computational account of the neural circuitry structuring it (fig. 1c). But the truth of this assumption is questionable, for there is no reason to think that the topographical boundaries of a brain area will always map onto those postulated by the best computational characterization of its neural circuitry (we also mentioned this point in reference to the AI in section 4.1). Consider, again, the case of the hippocampus. One of the main reasons to include the six regions mentioned above as comprising the hippocampal formation was the discovery of two unidirectional neuronal pathways (Andersen et al. [1971]): a trisynaptic pathway, going from EC to DG (synapse 1), from DG to CA3 (synapse 2), and then from CA3 to CA1 (synapse 3); and a monosynaptic pathway, directly connecting the EC

[9] An anonymous reviewer suggested a different alternative: that instead of thinking of a computational characterization in abstract labels that simply reuse terms from our current ontology, we should think that the right ontology for the computational operations would have as its components 'computational operations, which may or may not be expressible in human natural language terms'. This is an intriguing possibility, no doubt, but at this stage it is very unclear how it could possibility be implemented in the brain. For one, if the functions are non-expressible in natural human language terms, how would they inform psychological theories that are couched in human natural language terms? Second, as mentioned above, the most formal computational theories we have are highly task-dependent, as the values that the computational parameters can take must be specified over a particular numerical domain. Generalizing a single computational parameter over a large set of task domains is not a trivial matter. Thus, while interesting, we suspect that to fully evaluate the viability of this proposal may require its own paper.

with CA3 and DG via the perforant pathway. Indeed, the difference between these two pathways is the backbone of what's likely the most powerful computational model of two kinds statistical learning associated with the hippocampus (Shapiro et al. [2017]). The problem, though, is that recent evidence suggests that the old neuroanatomical mapping of the trisynaptic pathway is incomplete, and that its correct neuroanatomical characterization should include further projections to the subiculum and to other cortical areas not traditionally included in the hippocampus (Knierim [2015]). Given this, if the two-paths model for statistical learning is correct, then it is likely that a full understanding of how the hippocampus carries out such computations will require us to move beyond its anatomical limits. Indeed, it may turn out to be a mistake to attribute a computational function to the hippocampus, for the best way to characterize its operation-function might not respect its topographical boundaries at all.

4.3. The task-process barrier

In this final section, we challenge the way the top level, that is, the level of the psychological or behavioural functions ($P_1,\ldots,P_n$ in fig. 1), is characterized by the bottom-up approach. The starting point is the recognition that, contrary to what the bottom-up approach may suggest (fig. 1c), we never measure cognitive processes or behavioural functions directly (Francken et al. [2022]). What we measure is performance in tasks that we take to index behavioural processes or functions (fig. 1f). For instance, researchers may use the Stroop task to measure cognitive control, a false-belief task to study theory of mind, or an odd-ball task to measure attention. Unfortunately, often there is disagreement as to whether an experimental task actually measures the intended behavioural function, and sometimes these disagreements lead researchers to change their mind and accept that either the task does not index the intended behavioural function or that it actually measures a different one (fig. 1f).

Consider two examples: The first concerns the serial reaction time (SRT) task, which is widely used to measure motor learning (Nissen and Bullemer [1987]). In the SRT task, participants are visually presented with a horizontal arrangement of four dots that can turn on and off. They are also presented with a response box consisting of four buttons, also arranged horizontally, each one corresponding to a different finger. Sequences of visually presented cues in the form of 'on' dots on the screen are presented and participants are simply asked to respond in a spatially congruent manner in the button box (that is, leftmost dot corresponds to leftmost finger, second dot to the second finger, and so on). Both accuracy and reaction times are measured, and it is normally thought that increased accuracy and reduced reaction times reflect motor learning. However, recently some researchers have argued that the SRT task is not a good way to measure motor learning, for three reasons (Krakauer et al. [2019]). The first is that it can't distinguish between two arguably essential components of motor learning: reaction times and movement time. The second is that SRT measures accuracy as a percentage of correct responses, but says nothing about the quality of the execution of an action, which is a critical component of motor learning. And the third reason is that, contrary to the assumptions underlying the SRT task, learning does not occur implicitly but explicitly, as participants' awareness of the sequence seems to account for most of their successful performance. As a result, they argue that the SRT task should not be seen as measure of motor learning; at best, it measures the explicit learning of ordered sequences.

The second example concerns the *n*-back task, which is widely used to measure working memory (Kirchner [1958]). This task requires participants to monitor a sequence of stimuli and indicate when the current stimulus matches the one presented *n* steps back in the sequence. The thought is that as the number of steps back increases—that is, the larger *n* is—the more taxing the task is for working memory. Despite being widely used, though, many have argued that the *n*-back is not a reliable measure of working memory. Miller et al. ([2009]), for instance, assessed the reliability of the *n*-back task against other working memory tasks employed in neuropsychological testing, and found that they had very little convergent validity. The results, perhaps more worryingly, show no correlation between *n*-back task performance and backward digit spam recall, another widely used measure of working memory. More recently, Rac-Lubashevsky and Kessler ([2016]) studied individual differences in *n*-back task performance, and they identified two clear categories of performance: one corresponding to 'maintenance' and another one to 'updating'. These categories led them to suggest that the *n*-back task does not measure a single cognitive process but, likely, two. Lastly, Beukers et al. ([2024]) recently constructed a neural network model to simulate human performance in an *n*-back task and demonstrated that a working memory component alone cannot account for the retention of the maintained information, requiring instead a long-term episodic memory component to reach performance. They suggest that the *n*-back task, far from simply relying on working memory, may actually be tapping into episodic memory instead.

The fact that these disagreements occur with two of the most widely used tasks in cognitive neuroscience is actually diagnostic of a much more pervasive phenomenon—that the link from task performance to behavioural and/or psychological function is never straightforward, but depends on the experimenters' assumptions and background theoretical beliefs (Cronbach and Meehl [1955]). Many papers are rejected in the review process because the tasks do not clearly measure what they say they do, and the validity of experimental tasks as reliable measures of intended cognitive processes in many published papers is constantly questioned in laboratory meetings and professional conferences around the world. In sum, we often re-conceive what a task measures, not because something changed about the brain or the task, but because we often change how we understand the cognitive or behavioural function it is supposed to measure or the relationship between the task and its intended target. The result is that if we follow the bottom-up approach, we may end up mischaracterizing latent operation-functions associated with a particular task, but not because we select the wrong region. It would happen instead because we wrongly categorize the behavioural function it supposedly indexes. Big data may reduce this concern but cannot eliminate it. These local conceptual confusions won't get averaged out from aggregating massive amounts of brain data simply because they are not noise—they are just the wrong signal.

5. Conclusion: The Need for a Piece-Meal Approach

In the past decade, some researchers in the cognitive neurosciences have proposed employing big-data approaches to resolve disputes about structure–function mappings and cognitive ontologies (McCaffrey and Wright [2022]). In a recent proposal, Genon et al. ([2018]) offer a concrete strategy to do just that. Here we argued that their bottom-up approach suffers from serious shortcomings. Specifically, we argued that what constitutes a basic brain region is not obvious, and that its identification is likely never entirely *a priori*. We also argued that the characterization of the latent operation-function (that is, computation) of a given brain region is a more complex process than the bottom-up approach assumes. Finally, we argued that the bottom-up approach misses a critical conceptual barrier between task measurements and the intended target of the measure, posing serious challenges to the assumed straightforwardness with which big data is supposed to help to characterize the behavioural profile of a brain region.

Problems with big-data approaches to cognitive ontologies and structure–function mappings suggest that the fundamental difficulties with characterizing the function of brain regions go beyond the limits of experimental methods or the availability of empirical evidence. The solution very likely requires us to conceptually clarify, in advance, what are the right categories according to which brain data ought to be interpreted. Surely more data is better than less data (we are not denying that!), and it is very likely that better inferential statistics on big data repositories, such as those afforded by data science and machine learning approaches (for example, standardization, regulation via penalizing, and so on; see Rokem and Yarkoni [2023]), are going to be required to make better sense of the massive amounts of neuroscientific results labs around the world are producing daily. Nevertheless, it is very likely that we will always have to resolve local issues about the nature of brain regions, the adequacy of one or another computational account, or the convergent validity of one or another experimental task. This careful and conscientious 'piecemeal' approach to cognitive ontologies and structure–function mappings, which has been around for centuries, is likely not going to be replaced by solely top-down or bottom-up approaches any time soon, regardless of whether these approaches use the seductive power of big data.

Acknowledgements

Previous versions of this article were presented at the Philosophy Departments at Florida State University, Duke University, Mississippi State University, and the University of Cincinnati, as well as at the Symposia 'Philosophy, Behavior and the Brain' at the University of Virginia, and 'Fitting the mind to the brain: Deliberations on cognitive ontology' at Washington University in St. Louis. We are thankful for the comments from all those audiences. Special thanks, too, to Matthew Stanley, Trey Boone, and other members of the Imagination and Modal Cognition Lab, at Duke University, as well as two anonymous reviewers for their feedback.

Felipe De Brigard
Department of Philosophy
and
Department of Psychology and Neuroscience
and
Duke Institute for Brain Sciences
and
Center for Cognitive Neuroscience
Duke University
Durham, NC, USA
felipe.debrigard@duke.edu

Bryce S. Gessell
Department of Philosophy
Southern Virginia University
Buena Vista, VA, USA
bryce.gessell@svu.edu

References

Amaral, D. G., Scharfman, H. E. and Lavenex, P. [2007]: ‘The Dentate Gyrus: Fundamental Neuroanatomical Organization (Dentate Gyrus for Dummies)’, *Progress in Brain Research*, **163**, pp. 3–22, 788–90.

Andersen, P., Bliss, T. V. P. and Skrede, K. K. [1971]: ‘Lamellar Organization of Hippocampal Excitatory Pathways’, *Experimental Brain Research*, **13**, pp. 222–38.

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

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

Anderson, M. L., Kinnison, J. and Pessoa, L. [2013]: ‘Describing Functional Diversity of Brain Regions and Brain Networks’, *Neuroimage*, **73**, pp. 50–58.

Barack, D. L. [2019]: ‘Cognitive Recycling’, *British Journal for the Philosophy of Science*, **70**, pp. 239–68.

Bauernfeind, A. L., de Sousa, A. A., Avasthi, T., Dobson, S. D., Raghanti, M. A., Lewandowski, A. H., et al. [2013]: ‘A Volumetric Comparison of the Insular Cortex and Its Subregions in Primates’, *Journal of Human Evolution*, **64**, pp. 263–79.

Beukers, A. O., Hamin, M., Norman, K. A. and Cohen, J. D. [2024]: ‘When Working Memory May Be Just Working, Not Memory’, *Psychological Review*, **131**, pp. 563–77.

Bilder, R. M., Sabb, F. W., Cannon, T. D., London, E. D., Jentsch, J. D., Parker, D. S., Poldrack, R. A., Evans, C., Freimer, N. B. et al. [2009]: ‘Phenomics: The Systematic Study of Phenotypes on a Genome-Wide Scale’, *Neuroscience*, **164**, pp. 30–42.

Bohland, J. W., Bokil, H., Allen, C. B. and Mitra, P. P. [2009]: ‘The Brain Atlas Concordance Problem: Quantitative Comparison of Anatomical Parcellations’, *PLOS One*, **4**, available at <doi.org/10.1371/journal.pone.0007200>.

Boone, T. and De Brigard, F. [forthcoming]: ‘Philosophy of Cognitive Neuroscience’, in E. Craig (*ed*.), *Routledge Encyclopedia of Philosophy.*

Brockhaus, H. [1940]: ‘Cyto- and Myelo-architectonics of the Cortex Claustralis and the Cluastrum in Humans’, *Journal fur Psychologie und Neurologie*, **49**, pp. 249–348.

Brodmann, K. [1909]: *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*, Leipzig: J. A. Barth.

Burnston, D. C. [2016]: ‘Computational Neuroscience and Localized Neural Function’, *Synthese*, **193**, pp. 3741–62

Caramazza, A. [1986]: ‘On Drawing Inferences about the Structure of Normal Cognitive Systems from the Analysis of Patterns of Impaired Performance: The Case for Single-Patient Studies’, *Brain and Cognition*, **5**, pp. 41–66.

Coltheart, M. [2006]: ‘What Has Functional Neuroimaging Told Us about the Mind (so Far)?’, *Cortex*, **42**, pp. 323–31.

Cronbach, L. J. and Meehl, P. E. [1955]: ‘Construct Validity in Psychological Tests’, *Psychological Bulletin*, **52**, pp. 281–302.

De Brigard, F. [2017]: ‘Cognitive Systems and the Changing Brain’, *Philosophical Explorations*, **20**, pp. 224–41.

De Brigard, F. [2019]: ‘Know-How, Intellectualism, and Memory Systems’, *Philosophical Psychology*, **32**, pp. 719–58.

De Brigard, F. and Sinnott-Armstrong, W. [2022]: *Neuroscience and Philosophy*, Cambridge, MA: MIT Press.

Deen, B., Pitskel, N. B. and Pelphrey, K. A. [2011]: 'Three Systems of Insular Functional Connectivity Identified with Cluster Analysis', *Cerebral cortex*, **21**, pp. 1498–506.

Evrard, H. C., Logothetis, N. K. and Craig, A. D. [2014]: 'Modular Architectonic Organization of the Insula in the Macaque Monkey', *Journal of Comparative Neurology*, **522**, pp. 64–97.

Evrard, H. C. [2019]: 'The Organization of the Primate Insular Cortex', *Frontiers in Neuroanatomy*, **13**, available at <doi.org/10.3389/fnana.2019.00043>.

Fodor, J. [1999]: 'Why the Brain?', *London Review of Books*, **21**, available at <www.lrb.co.uk/the-paper/v21/n19/jerry-fodor/diary>.

Fox, P. T. and Lancaster, J. L. [2002]: 'Mapping Context and Content: The BrainMap Model', *Nature Reviews Neuroscience,* **3**, pp. 319–21.

Fox, P. T. and Friston, K. J. [2012]: 'Distributed Processing; Distributed Functions?', *Neuroimage*, **61**, pp. 407–26.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. and Dolan, R. J. [1996]: 'The Trouble with Cognitive Subtraction', *Neuroimage*, **4**, pp. 97–104.

Francken, J. C., Slors, M. and Craver, C. F. [2022]: 'Cognitive Ontology and the Search for Neural Mechanisms: Three Foundational Problems', *Synthese*, **200**, available at <doi.org/10.1007/s11229-022-03701-2>.

Gall, F. J. and Spurzheim, J. G. [1810]: *Anatomie et physiologie du système nerveux en général, et du cerveau en particulier, avec des observations sur la possibilité de reconnaître plusieurs dispositions intellectuelles et morales de l'homme et des animaux, par la configuration de leurs têtes*, Paris: Schoell.

Genon, S., Reid, A., Langner, R., Amunts, K., Eickhoff, S. B. and Schilbach, L. [2018]: 'How to Characterize the Function of a Brain Region', *Trends in Cognitive Sciences*, **22**, pp. 350–64.

Gessell, B., Geib, B. and De Brigard, F. [2021]: 'Multivariate Pattern Analysis and the Search for Neural Representations', *Synthese*, **199**, pp. 12869–89.

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V., Nichols, T. E. and Poldrack, R. A. [2015]: 'NeuroVault.org: A Web-Based Repository for Collecting and Sharing Unthresholded Statistical Maps of the Human Brain', *Journal of Neuroscience Methods*, **229**, pp. 8–14.

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M. and Petersen, S. E. [2016]: 'Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations', *Cerebral Cortex*, **26**, pp. 288–303.

Haueis, P. [2014]: 'Meeting the Brain on Its Own Terms', *Frontiers in Human Neuroscience*, 8, available at <doi.org/10.3389/fnhum.2014.00815>.

Herculano-Houzel, S. [2009]: ‘The Human Brain in Numbers: A Linearly Scaled-up Primate Brain’, *Frontiers in Human Neuroscience*, **31**, available at <doi.org/10.3389/neuro.09.031.2009>.

Huettel, S. A., Song, A. W. and McCarthy, G. [2014]: *Functional Magnetic Resonance Imaging*, Sunderland, MA: Sinauer Associates.

Huster, R. J., Debener, S., Eichele, T. and Herrmann, C. S. [2012]: ‘Methods for Simultaneous EEG-fMRI: An Introductory Review’, *Journal of Neuroscience*, **32**, pp. 6053–60.

Kanai, R. and Rees, G. [2011]: ‘The Structural Basis of Inter-individual Differences in Human Behaviour and Cognition’, *Nature Reviews Neuroscience*, **12**, pp. 231–42.

Kashyap, S., Ivanov, D., Havlicek, M., Poser, B. A. and Uludağ, K. [2018]: ‘Impact of Acquisition and Analysis Strategies on Cortical Depth-Dependent fMRI’, *Neuroimage*, **168**, pp. 332–44.

Kim, E. J., Sidhu, M., Gaus, S. E., Huang, E. J., Hof, P. R., Miller, B. L. et al. [2012]: ‘Selective Frontoinsular von Economo Neuron and Fork Cell Loss in Early Behavioral Variant Frontotemporal Dementia’, *Cerebral Cortex*, **22**, pp. 251–59.

Kirchner, W. K. [1958]: ‘Age Differences in Short-Term Retention of Rapidly Changing Information’, *Journal of Experimental Psychology*, **55**, pp. 352–58.

Klein, C. [2012]: ‘Cognitive Ontology and Region- versus Network-Oriented Analyses’, *Philosophy of Science,* **79**, pp. 952–60

Knierim, J. J. [2015]: ‘The Hippocampus’, *Current Biology*, **25**, pp. R1116–21.

Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L. and Haith, A. M. [2019]: ‘Motor Learning’, *Compr Physiol*, **9**, pp. 613–63.

Krockenberger, M., Saleh, T. O., Logothetis, N. K. and Evrard, H. C. [[unpublished]2020]: ‘Connection ‘Stripes’ in the Primate Insula’, available at <doi.org/10.1101/2020.11.03.361055>.

Mah, Y. H., Husain, M., Rees, G. and Nachev, P. [2014]: ‘Human Brain Lesion-Deficit Inference Remapped’, *Brain*, **137**, pp. 2522–31.

Marr, D. [1971]: ‘Simple Memory: A Theory of Archicortex’, *Philosophical Transactions of the Royal Society of London*, **262**, pp. 23–81.

McCaffrey, J. and Wright, J. [2022]: ‘Neuroscience and Cognitive Ontology: A Case for Pluralism’, in F. De Brigard and W. Sinnott-Armstrong (*eds*), *Neuroscience and Philosophy*, Cambridge, MA: MIT Press, pp. 427–66.

McCaffrey, J. B. [2023]: ‘Evolving Concepts of Functional Localization’, *Philosophy Compass*, **18**, available at <doi.org/10.1111/phc3.12914>.

McHugh, T. J., Jones, M. W., Quinn, J. J., Balthasar, N., Coppari, R., Elmquist, J. K. et al. [2007]: 'Dentate Gyrus NMDA Receptors Mediate Rapid Pattern Separation in the Hippocampal Network', *Science*, **317**, pp. 94–99.

Miller, K. M., Price, C. C., Okun, M. S., Montijo, H. and Bowers, D. [2009]: 'Is the *N*-back Task a Valid Neuropsychological Measure for Assessing Working Memory?', *Archives of Clinical Neuropsychology*, **24**, pp. 711–17.

Morris, R. [2007]: 'Theories of Hippocampal Function', in P. Andersen, R. Morris, D. Amaral, T. Bliss and J. O'Keefe (*eds*), *The Hippocampus Book*, Oxford: Oxford University Press, pp. 581–714.

Moustafa, A. A. [2017]: *Computational Models of Brain and Behavior*, Hoboken, NJ: John Wiley.

Nissen, M. J. and Bullemer, P. [1987]: 'Attentional Requirements of Learning: Evidence from Performance Measures', *Cognitive Psychology*, **19**, pp. 1–32.

O'Keefe, J. and Dostrovsky, J. [1971]: 'The Hippocampus as a Spatial Map: Preliminary Evidence from Unit Activity in the Freely-Moving Rat', *Brain Research*, **34**, pp. 171–75.

O'Keefe, J. and Nadel, L. [1978]: *The Hippocampus as a Cognitive Map*, Oxford: Oxford University Press.

Poldrack, R. A. and Yarkoni, T. [2016]: 'From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure', *Annual Review of Psychology*, **67**, pp. 587–612.

Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D. S., Sabb, F. W., Bilder, R. M. et al. [2009]: 'The Cognitive Atlas: Toward a Knowledge Foundation for Cognitive Neuroscience', *Frontiers in Neuroinformatics*, **3**, available at <doi.org/10.3389/fninf.2011.00017>.

Poldrack, R. A., Laumann, T. O., Koyejo, O., Gregory, B., Hover, A., Chen, M. Y., Gorgolewski, K. J., Luci, J., Joo, S. J., Boyd, R. L. et al. [2015]: 'Long-Term Neural and Physiological Phenotyping of a Single Human', *Nature Communications*, **6**, pp. 1–10.

Poldrack, R. A. and Yarkoni, T. [2016]: 'From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure', *Annual Review of Psychology*, **67**, pp. 587–612.

Posner, M. I. and DiGirolamo, G. J. [2000]: 'Cognitive Neuroscience: Origins and Promise', *Psychological Bulletin*, **126**, pp. 873–89.

Price, C. J. and Friston, K. J. [2002]: 'Degeneracy and Cognitive Anatomy', *Trends in Cognitive Sciences*, **6**, pp. 416–21.

Price, C. J. and Friston, K. J. [2005]: 'Functional Ontologies for Cognition: The Systematic Definition of Structure and Function', *Cognitive Neuropsychology*, **22**, pp. 262–75.

Rac-Lubashevsky, R. and Kessler, Y. [2016]: 'Decomposing the *N*-back Task: An Individual Differences Study Using the Reference-Back Paradigm', *Neuropsychologia*, **90**, pp. 190–99.

Raichle, M. E. [1983]: 'Positron Emission Tomography', *Annual Review of Neuroscience*, **6**, pp. 249–67.

Rokem, A. and Yarkoni, T. [2023]: *Data Science for Neuroimaging: An Introduction*, Princeton, NJ: Princeton University Press.

Schultz, W., Dayan, P. and Montague, P. R. [1997]: 'A Neural Substrate of Prediction and Reward', *Science*, **275**, pp. 1593–99.

Shallice, T. [1988]: *From Neuropsychology to Mental Structure*, Cambridge: Cambridge University Press.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M. and Norman, K. A. [2017]: 'Complementary Learning Systems within the Hippocampus: A Neural Network Modelling Approach to Reconciling Episodic Memory with Statistical Learning', *Philosophical Transactions of the Royal Society B*, **372**, available at <doi.org/10.1098/rstb.2016.0049>.

Shine, J. M., Eisenberg, I. and Poldrack, R. A. [2016]: 'Computational Specificity in the Human Brain', *Behavioral and Brain Sciences*, **39**, available at <doi.org/10.1017/S0140525X1500165X>.

Stachenfeld, K. L., Botvinick, M. M. and Gershman, S. J. [2017]: 'The Hippocampus as a Predictive Map', *Nature Neuroscience*, **20**, pp. 1643–53.

Sternberg, S. [1969]: 'The Discovery of Processing Stages: Extensions of Donders' Method', *Acta Psychologica*, **30**, pp. 276–315.

Tremblay, P. and Dick, A. S. [2016]: 'Broca and Wernicke Are Dead, or Moving Past the Classic Model of Language Neurobiology', *Brain and Language*, **162**, pp. 60–71.

Uttal, W. R. [2001]: *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*, Cambridge, MA: MIT Press.

Uttal, W. R. [2011]: *Mind and Brain: A Critical Appraisal of Cognitive Neuroscience*, Cambridge, MA: MIT Press.

Uttal, W. R. [2012]: 'Reliability in Cognitive Neuroscience: A Meta-meta-analysis', *Perspectives on Psychological Science*, **7**, pp. 632–42.

Viola, M. [2017]: 'Carving Mind at Brain's Joints: The Debate on Cognitive Ontology', *Phenomenology and Mind*, **12**, pp. 162–72.

Vogt, O. [1911]: ‘Die Myeloarchitektonik des Isocortex Parietalis’, *Journal für Psychologie und Neurologie*, **18**, pp. 379–90.

Ward, Z. B. [2022]: ‘Registration Pluralism and the Cartographic Approach to Data Aggregation across Brains’, *British Journal for the Philosophy of Science*, **73**, pp. 47–72.

Westfall, J. and Yarkoni, T. [2016]: ‘Statistically Controlling for Confounding Constructs Is Harder Than You Think’, *PLOS One*, **11**, available at <doi.org/10.1371/journal.pone.0152719>.

Whittington, J. C., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N. and Behrens, T. E. [2020]: ‘The Tolman–Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation’, *Cell*, **183**, pp. 1249–63.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. and Wager, T. D. [2011]: ‘Large-Scale Automated Synthesis of Human Functional Neuroimaging Data’, *Nature Methods*, **8**, pp. 665–70.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R. et al. [2011]: ‘The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity’, *Journal of Neurophysiology*, **106**, pp. 1125–65.

Zipser, D. [1985]: ‘A Computational Model of Hippocampal Place Fields’, *Behavioral Neuroscience*, **99**, pp. 1006–18.