Neural Reuse and the Nature of Evolutionary Constraints[1]

Charles Rathkopf[2]

Abstract. In humans, the reuse of neural structure is particularly pronounced at short, task-relevant timescales. Here, an argument is developed for the claim that facts about neural reuse at task-relevant timescales conflict with at least one characterization of neural reuse at an evolutionary timescale. It is then argued that, in order to resolve the conflict, we must conceptualize evolutionary-scale reuse more abstractly than has been generally recognized. The final section of the paper explores the relationship between neural reuse and human nature. It is argued that neural reuse is not well-described as a process that constrains our present cognitive capacities. Instead, it liberates those capacities from the ancestral tethers that might otherwise have constrained them.[3]

[1] This is a pre-print. The published version will appear in:
Calzavarini, F. & Viola, M. (Eds.). (2020). Neural mechanisms: New challenges in the philosophy of neuroscience. Springer.

[2] Forschungszentrum Jülich. Institute for Neuroscience and Medicine.

[3] Thanks to Matteo Colombo, Philipp Haueis, and Lena Kästner for their insightful feedback on my *Neural Mechanisms Online* talk, which was my first attempt to work out the issues discussed in this chapter.

1 A latent disagreement about neural reuse

One might think that each time an organism acquires a novel behavioral capacity, some correspondingly novel structure must have been wired together in its head. Neural reuse is the contrasting idea that novel capacities are often made possible by the redeployment of existing neural structures in new task domains. Here, I hope to identify a latent disagreement in the scientific discussion of neural reuse.

The disagreement has remained latent because it concerns the relationship between two background assumptions, which have themselves received little attention. The first assumption concerns the multiplicity of timescales at which neural reuse might occur. The second concerns the role of representation in theories of neural function. These two topics come together in a particularly interesting way in Stanislas Dehaene's work on reading acquisition. After introducing neural reuse more thoroughly, I will give a brief overview of Dehaene's theory, and draw from it a principle about how timescale and representational character are related. That principle – which I call the content constraint view – is not the only way to conceive of the relationship between timescale and representational character. I sketch an alternative view of this relationship, and then work out three consequences of accepting that alternative view, each of which serves to refine our understanding of neural reuse.

In the final section of the paper, I explore a loftier and more speculative set of ideas about the relationship between neural reuse and human nature. It is argued that, if the view of neural reuse developed earlier in the paper is right, then neural reuse helps explain how human nature managed to acquire its uniquely open-ended character.

2 Reuse: a central theme, and its variations

Here, I use the term "reuse" in a maximally broad sense, intended to capture a common theme running through a complex and partially overlapping set of theories. Labels for these theories include "neural repurposing" (Parkinson and Wheatley, 2015), "neuronal recycling" (Dehaene and Cohen, 2007), "massive redeployment" (Anderson, 2007), "cognitive recycling" (Barack, 2017), and "neural exaptation" (Chapman et al., 2017). Neural reuse, in the maximally broad sense intended here, is entailed by each theory in this list. It can be defined as a commitment to two simple ideas. The first is that local neural structures contribute to multiple cognitive or behavioral tasks. The term "local neural structure" is meant to be quite inclusive. It covers everything from cytologically-defined microscale structures, such as cortical columns, all the way up to functionally defined cortical regions identified by means of brain imaging.

The second idea is that the cognitive or behavioral tasks to which a structure contributes must be conceptually distinct. If the latter function logically entails the former, the two functions are not conceptually distinct. A non-scientific example may be helpful here. Consider the following two claims. On Monday, my travel mug is used to transport coffee. On Tuesday, it is used to transport hot coffee. Because "transporting hot coffee" entails "transporting coffee," this is not a case of reuse in the relevant sense. To make this a case of reuse in the relevant sense, I would have to transport something conceptually unrelated, like soup.[4] Now let's consider a neuroscientific example. In task condition A, the supplementary motor area (SMA) subserves motor command preparation. In task condition B, the SMA subserves reaching movement preparation. Because preparation for a reaching movement is one kind of motor command preparation, these two functions are not conceptually distinct. The conceptual overlap between these two functions blurs the distinction between the theory of neural reuse and the comparatively bland claim that neural function is subject to variation of some sort or another. In a review paper on the SMA that focuses on conceptual difficulties associated with theories of SMA function, Naschev et al. put the point thus: "Functional pleomorphism is conceptually problematic owing to the difficulty of explaining the process of switching between different neural functions" (Naschev et al., 2008). Another function sometimes ascribed to the SMA is the regulation of task-switching, which is arguably distinct from movement preparation, and would, therefore, support the case for neural reuse in that area.

The dual characterization provided thus far shows what the various theories of neural reuse have in common. They differ from one another in many dimensions, two of which are relevant here. The first has to do with timescale. What are the timescales at which neural reuse occurs? A view that is commonly assumed, if not explicitly defended, is that there are exactly two such scales: one phylogenetic and one ontogenetic (Gallese, 2008; Anderson and Finlay, 2014). Such an assumption appears to be held, for example, by Parkinson and Wheatley (2015), who divide their discussion of the topic into "neural repurposing across lifetimes" and "neural repurposing within lifetimes." It is also commonly assumed, if not explicitly defended, that the reuse process at the phylogenetic scale stands in a relatively harmonious relationship to reuse at the ontogenetic scale. At the very least, none of the existing literature explores the possibility that our description of neural reuse at one scale will carry implications for the viability of description at another. This assumption can be challenged. As I argue below, once we explore the possibility of

[4] In this prosaic example, there is no deep truth about which functions are genuinely distinct, because the individuation conditions for the functions of a coffee mug are, presumably, a matter of convention rather than discovery.

additional timescales, the relations between these two default scales begin to look less harmonious.

Another dimension of difference between theories of neural reuse concerns the kinds of purposes, or functions, that a theory might describe at each scale. Even after we have restricted ourselves to a single scale in space and time, the varieties of neural function are many. Some functions are characterized in terms of proximate effects on other neural structures; others in terms of distal effects on behavior. Functions can also be distinguished with respect to the faculty to which they contribute: perception, memory, motor control, etc. The distinction I want to draw, which I take to be orthogonal both to the proximal/distal distinction, and to the choice of mental faculty, divides what I will call content functions from all others. A content function is any function in which the contribution a structure makes to the operation of the system of which it is a part involves the representation of an element in the task-environment of the organism.

Two components of this definition deserve some unpacking. The first is the concept of a neural representation. In most areas of neuroscience, the term "representation" is used liberally.[5] The concept I mean to invoke here has a more distinctive theoretical role. A pattern of activity only counts as a representation, in the sense I have in mind, if (i) it is correlated with some environmental parameter of relevance, and (ii) it plays a causal role in the cognitive process that enables the organism to achieve some behavioral goal, by acting as a signal that informs the activities of downstream neural mechanisms. This account of representation is incomplete, but useful. The first condition suffices to rule out neural activity that systematically influences behavior without targeting external properties. The second condition rules out what have been called idle correlations (Rathkopf, 2017), which fail to figure in the representational activities of the organism because no mechanism exists that is capable of exploiting the correlation in order to direct behavior.

The second component in the definition of content function that deserves unpacking is the concept evoked by the phrase "element in the task-environment of the organism." To be an element in the task-environment of the organism is to be the kind of property to which the organism must at some point dedicate attention, in order to complete a particular task successfully. Consider, for example, the so-called fusiform face area (FFA) in humans. It has been described as cortical structure that is dedicated to the detection of faces (Kanwisher, 2010). The representations of faces purportedly instantiated by that structure must be consulted before one can, for example, appropriately orient one's gaze toward a

[5] To see this, consider how difficult it is to design an experiment that might serve to falsify the claim that "x is a representation," where x is any pattern of neural activity you choose.

conversational partner. Faces, therefore, will commonly count as elements in the task environment of humans, and face-detection will commonly count as a content function.

The class of non-content functions will include both neural functions that do not demand representational characterization, along with neural functions that do, but which are only indirectly connected with what would ordinarily be countenanced as a task. As Phillip Haueis (2018) has recently argued, there are many kinds of representational activity in the brain that are only indirectly involved with the accomplishment of intuitively recognizable behavioral goals, and which, therefore, have only a tenuous connection to familiar, folk-psychological modes of description. Moreover, there are many neural activities that play roles that are both highly specific and vital to the life of the organism, but which do not admit of representational description at all. Pacemaker neurons, for example, dampen the dynamics of various neural networks by means of intrinscally modulated bursting activity (Ramirez et al., 2004). Purkinje cells in the cerebellum have been described as gain modulators, that multiply incoming signals from a wide variety of perceptual sources (Luque et al., 2019). Cases like these remind us that neural reuse need not, as a matter of definition, consist exclusively in transitions between content functions.

Thus far, I have introduced a very general notion of neural reuse, and introduced two ways to distinguish between the many kinds of neural function that might be involved in any given case of neural reuse. First, I distinguished between neural functions instantiated on task-relevant time scale and those instantiated on an evolutionary time scale. Second, I distinguished between content functions and non-content functions. The core insight in this essay is that these two distinctions are empirically linked. If we characterize the function of a local neural structure at the timescale of an individual task, we may find good evidence that it realizes a content function. If, however, we try to characterize its function on larger timescales, we are likely to find that the evidence for content functions disappears. Before I present the argument that shows how timescale and representational status are related, it will be helpful to examine a particular theory of neural reuse and its application to a particular cognitive phenomenon. For this purpose, I have chosen Stanislas Dehaene's theory of neuronal recycling and its application to literacy. Dehaene's theory is appropriate for the job, not only because of the strength of its influence, which is considerable, but also because it illustrates the logic behind a view of the relationship between biological evolution and mental content that is implicit in a lot of evolutionary psychology, but which, I'll argue, ought to be resisted.

3 The Paradox of Reading

In his book “Reading in the brain," Dehaene presents a theory of reading and reading acquisition. The book begins by introducing what Dehaene calls the reading paradox, which is most succinctly expressed in the following two sentences: “Nothing in our evolution could have prepared us to absorb language through vision. Yet brain imaging demonstrates that the brain contains fixed circuitry exquisitely attuned to reading (Dehaene, 2009, p.24)." Dehaene’s version of neural reuse, which he calls the neuronal recycling hypothesis, is offered as a solution to this paradox. To understand his theory, then, we first need to understand this paradox in more detail, and some of the data that appear to generate it.

The reading paradox presents us with two claims that are, ostensibly, both true and mutually inconsistent. The first is about human evolution. We know from anthropological evidence that the earliest human writing systems appeared about 6,000 years ago, in the form of Mesopotamian cuneiform (d’Errico and Colagè, 2018). We also know from mutation frequency data that 6,000 years is too short a period for substantial neurogenetic adaptations to have accumulated. We can be confident, therefore, that the capacity for literacy is not the direct product of a genetic mutation that has only recently swept through the human gene pool.

The second half of the paradox also deserves a closer look. What does it mean to say that “the brain contains fixed circuitry, exquisitely attuned to reading?" The circuitry to which Dehaene refers is a small, functionally defined cortical area located in the left ventral occipito-temporal junction. That area is now commonly labeled with a functional designation that Dehaene himself coined: the visual word form area, or VWFA. Dehaene ascribes two properties to this circuitry. He says that it is fixed, and that it is exquisitely attuned to reading. Let us first examine what he means by the latter. Dehaene’s claim that the VWFA is exquisitely attuned to reading is what he takes to be the upshot of a family of interesting results from lesion and imaging data, which, when taken as a whole, suggest that, in literate adult subjects, the area is specialized for word recognition.

The following six pieces of evidence are commonly taken to provide support for this localizationist conclusion.

1. In normal literate subjects, the region is differentially responsive to written, but not spoken words (Dehaene and Cohen, 2007).

2. Illiterate adults do not show responsivity to letters in VWFA, and ex-illiterate adults (people who first learned to read in adulthood) exhibit less responsivity than literates. (Thiebaut et al., 2012).
3. In blind subjects, the region is differentially responsive to words presented in Braille, but not to tactile control stimuli (Reich et al., 2011).[6]
4. Lesions to the area appear to result in pure alexia, a condition in which formerly literate subjects cannot understand written words, despite being able to understand and produce verbal speech at roughly normal levels of competency (Gaillard et al., 2006).
5. fMRI priming effects in this region are invariant to alternative representations of the same priming word. For example, the stimulus "RADIO" is an effective prime for "radio," whereas "oidar" is not (Dehaene and Cohen, 2007).
6. The repetition suppression effect disappears in this region for mirror-images of words and individual letters. The visual system regards most objects as equivalent to their mirror-images. We learn to violate this rule when learning to read, in order to distinguish, for example, "b" from "d." That this region responds differently to mirror images suggests that the region is sensitive to words as meaningful units, rather than as linear strings of wiry objects (Dehaene and Dehaene-Lambertz, 2016; Dehaene, 2013).

These results provide strong evidence that the brains of literate adults contain an area with a response profile dominated by words and letters. If Dehaene's interpretation of the data is correct, then the overriding function of the VWFA is to represent words and letters. Since words and letters are elements of common human task environments, Dehaene's hypothesis describes a content function, in the sense defined above.

The apparently localized nature of word recognition is fascinating in its own right, but what exactly is its relevance to the paradox of reading? On Dehaene's view, it is a theoretical surprise that word recognition appears to be carried out in such a small and discrete cortical area. The sense of surprise is reinforced by the claim that this area is "fixed." This term refers to the fact that the spatial position of the area, despite being functionally rather than anatomically identified, is robust across individual subjects and language groups.[7] The combination of response-specificity and positional robustness characteristic of the VWFA is loosely analogous to the kinds of retinotopic maps found in early visual cortex. By analogy to areas like these, Dehaene expects that, in general, positionally robust, map-like

[6] Although this claim has recently been disputed, in light of new data. See Kim et al. (2017).
[7] Although see Coltheart (2014) for a somewhat deflationary interpretation of the degree of positional robustness that is actually licensed by the neuroimaging data.

circuits in human cortex will subserve capacities that emerged long ago and that are part of our biological, rather than cultural, heritage.

Now that we have a firmer grasp on the meaning of the two claims involved in the paradox of reading, we can ask: is it reasonable to characterize them as a paradox? Perhaps not. If we streamline the wording a bit, the purported paradox juxtaposes the claim that (i) orthographic word identification is a localized brain function, with the claim that (ii) orthographic word identification could not have played a role in human evolution. From a logical point of view, these claims are not actually inconsistent. If their conjunction appears paradoxical, it is only because we have tacitly accepted a background assumption which says that localized content functions are necessarily driven by the genetic evolution of the species.

Like many assumptions lurking in the scientific background, this one arouses suspicion as soon as it is formulated explicitly and offered up for critical inspection. The assumption asks us to contrast evolved functions with learned ones. But, as developmental systems theorists have emphasized, this contrast is easily abused, because every neural function emerges from a process of biological development, and the distinction between development and learning is both highly theoretical and highly contested (Oyama, 2000). Moreover, even on a thin conception of learning, there are no uncontroversial examples of content functions that develop in its absence. In light of the entangled nature of evolution and development, any theory that requires us to assign causal responsibility for a trait to one process or the other should at least be explicit about how the assignment should be carried out. Since the assumption in this case is merely implicit, no such instructions are provided. It is reasonable to suspect, therefore, that the conceptual foundations underlying the assumption are unstable. In Section 6, I'll argue that the assumption should be rejected. In the following section, however, we examine Dehaene's favored solution instead.

4 Neuronal Recycling as a Solution to the Paradox

Because Dehaene leaves untouched the assumption linking localization and evolutionary provenance, the only way he can solve the paradox of reading is by showing that, contrary to first appearance, one of the two claims that comprise the paradox is not strictly true. Dehaene aims to undermine, or at least weaken, the claim about evolution. The theory of neuronal recycling says that, although natural selection cannot be directly responsible for having shaped a circuit dedicated to reading, natural selection is, nevertheless, responsible for having indirectly shaped the mechanism that enables us to read. Natural selection shaped a circuit for a particular function that is sufficiently close to reading, but which, unlike reading itself, reaches far back into human evolutionary history.

> Cultural acquisitions (e.g., reading) must find their "neuronal niche," a set of circuits that are sufficiently close to the required function and sufficiently plastic as to reorient a significant fraction of their neural resources to this novel use (Dehaene and Cohen, 2007).

Here, and in other passages, Dehaene appeals to a principle of similarity between functions to explain what makes it the case that they share the same cortical fate. The similarity relation holds between an older function and a newer one. At this point, it will be useful to introduce a pair of terminological stipulations. In any case of neural reuse, whether it occurs on an evolutionary scale or not, I'll refer to the older function as the primary function, and the newer one as the secondary function. A core commitment of neuronal recycling can then be expressed as follows: primary functions are necessarily similar to secondary functions. When expressed this way, the obscurity of the claim looms large. Similarity with respect to what?

In Dehaene's 2009 book, as well as in many of the articles he has produced with various co-authors on the topic, including the 2007 article with Laurent Cohen, (from which the quote above is drawn) his answer to this question appears to be that the relevant kind of similarity is similarity with respect to content. Dehaene stresses that, according to neuronal recycling, cortical circuits are typically biased towards the representation of certain elements of the organism's task environment. These biases serve to constrain the range of cultural symbols humans can learn to use.

> According to this view, our evolutionary history, and therefore our genetic organization, specifies a cerebral architecture that is both constrained and partially plastic, and that delimits a space of learnable objects. New cultural acquisitions are possible only inasmuch as they are able to fit within the pre-existing constraints of our brain architecture (Dehaene, 2008, p.12).

What kinds of neural properties have the power to delimit the space of learnable objects, as Dehaene puts it? One might attempt to answer this question in terms of content-neutral limitations on the systems' capacity to process information. If the object is too complex for the perceptual system to discriminate, for example, it is not a learnable object. (This is, presumably, one reason that no written languages employ symbols with 1000 overlapping components.) However, this is not the kind of answer Dehaene has in mind. Dehaene's view seems to be that the limitation is neither merely perceptual, nor directly related to the complexity of the object. On Dehaene's view, we have an inherited "preference" for objects

with particular semantic qualities. These content preferences are genetically entrenched, and it is in virtue of that entrenchment that the space of learnable objects is limited. On this view, unless some very sophisticated genetic engineering becomes a viable option, the space of learnable objects is destined to remain circumscribed.

This focus on evolutionarily entrenched content is one way of making sense of two bodies of evidence. The first body of evidence is the response specificity of the VWFA, which was described above. The second body of evidence is the fact that all known written languages employ characters with specific geometric similarities. For example, if you plot the distribution of the number of line crossings required to represent all of the written characters in all of the world's languages, you get a tight cluster around the number three (Changizi and Shimojo, 2005). Dehaene also cites as evidence the (purported) fact that written characters in all human languages are necessarily composed of combinations of elementary shapes. Dehaene sees both bodies of evidence (response specificity and orthographic similarity across languages) as effects of a hidden common cause - the content bias in VWFA. The content bias is postulated, by means of an inference to the best explanation, precisely in order to account for both the neural and the anthropological data.[8]

To summarize the foregoing remarks, Dehaene's theory of neuronal recycling is offered as a solution to the paradox of reading. It counts as a solution because it purports to show that the evolutionary claim that constitutes the first half of the paradox is, despite its initial plausibility, wrong. Evolution did indeed "prepare us to absorb language through vision," but it did so indirectly. What I will *the content constraint view* is a theory about that process of indirect preparation. It can be split into two claims.

1. The primary evolutionary function of the VWFA is a content function.
2. Constraints on the range of secondary functions for which the VWFA can be "recycled" derive from the nature of the content targeted by its primary function.

In the following section, I provide reasons to think that the content constraint view is incorrect. In his most recent work on the topic, Dehaene et al. (2018) defend a view of the VWFA that is in tension with the content constraint view. One might worry, therefore, that I have been constructing a straw man. However, my motivation for articulating the view is not to weigh in on debates about the neural substrates of literacy. It is rather to articulate a

[8] The anthropological data Dehaene offers as evidence of neural reuse may be not as straightforward as he sometimes makes it sound. Max Coltheart has argued that the uniformity to which Dehaene refers is simply not there (Coltheart, 2014). I am sympathetic to Coltheart's concerns about the evidence, but would like to resist Dehaene's account on different grounds altogether. I will therefore just assume the evidence says exactly what Dehaene says it does.

conception of neural reuse in which content plays a central explanatory role, even on an evolutionary scale. The content constraint view is worth articulating not because it has arduous defenders who happen to be wrong, or because it has a severely detrimental effect on the design of new experiments, but because the consequences of rejecting it are theoretically interesting. Once we reject it, I'll argue, we see that theories of neural reuse, when pitched at an evolutionary scale, are more enigmatic than has been recognized thus far.

5 A Clash Between Timescales

The content constraint view describes a process that bridges two timescales. The primary function gets stabilized on an evolutionary timescale. It plays an important role in the selection history of the organism, and thereby leaves a trace on the genetic information transmitted across generations. That genetic information manifests itself in the form of a content bias, which is itself expressed by a particular local structure. The secondary function operates on a different timescale altogether. It gets stabilized on a developmental scale. The target of the secondary function is determined in part by developmental context and cultural input, but is also constrained by the content bias in the circuit that subserves it. In what follows, the target of my attention is the nature of this purported constraint, and how it might have come about over evolutionary time.

The challenge I want to pose emerges from thinking about the evolutionary implications of another kind of neural reuse; one that unfolds more quickly than the kind Dehaene describes. This faster process, which I call task-scale neural reuse, is a phenomenon in which a local neural structure transitions from supporting one behavioral task to supporting another by means of a reconfiguration of its network of partnering structures. Such reconfiguration unfolds on a timescale relevant to individual cognitive and behavioral tasks, on the order of seconds or minutes. On this view, each structure supports different functions at different times, depending not only on the current perceptual input, but also on set of structures with which functional connectivity has been established.

The evidence for this architectural principle is multifaceted. One of the more significant sources of evidence comes from meta-analyses of brain imaging studies on humans. For example, Anderson et al. (2013) ask how many distinct tasks, drawn from distinct cognitive domains, are supported by each region of the brain. To estimate an answer to this question, they measure voxel-by-voxel diversity in data generated by a collection of over 2,000 functional neuroimaging experiments. The analysis shows that even small regions of the brain contribute to multiple tasks both within and between cognitive domains (Anderson et al., 2013).

> The upshot: local neural structures are not highly selective and typically contribute to multiple tasks across domain boundaries. Because the domains are highly varied, the observations cannot be explained by the similarity of the task domains (Anderson, 2014, p.10).

This passage is particularly appropriate for our exposition of Anderson's view because it is explicit about the absence of an underlying similarity relation that could serve to unify or circumscribe the set of tasks that a given structure, could, in principle, be recruited to support. If the list of functions associated with each structure ranges across both tasks and cognitive domains, then no structure specializes in the representation of a particular element in a particular task-environment. In other words, no structure specializes in any particular content function. The anti-localizationist implications of task-scale neural reuse are well known, and detailed arguments to this effect can be found elsewhere (Bergeron, 2010; Rathkopf, 2013; McCaffrey, 2015; Zerilli, 2019).

There is also reason to believe that the distributed functional architecture implied by task-scale neural reuse has always been a feature of the human brain. Macaque cortex, for example, appears to implement a form of task-scale neural reuse (Iriki and Taoka, 2012), and the last common ancestor of macaques and humans lived approximately 25 million years ago (Distoll and Tosi, 2007). The idea that task-scale neural reuse is ancient in our lineage poses a direct threat to the content constraint view. To see this, we need only ask what justification we have for claiming that some neural structure has a primary function that can be characterized in terms of content. Typically, the biological justification for isolating one primary function from the myriad causal interactions in which a given structure may be engaged involves an appeal to natural selection. But if task-scale neural reuse is ancient, natural selection will have had little opportunity to tailor a structure for its capacity to contribute to any particular content function.

This argument shows that if we want to characterize the contribution of a neural structure to the capacities of an organism on an evolutionary scale, we cannot invoke any particular content-function. And this claim, in turn, conflicts with the content constraint view. If the evolution of local neural structures was not driven by the demands of dealing with particular kinds of content, then constraints on the range of secondary functions that those structures can come to realize are not accurately described as constraints on content. Of course, this argument does not show that the range of secondary functions a neural structure can come to support is *unconstrained*. Nor does it show that the operative constraints, whatever they are, are not bound up with the evolutionary history of the organism. It only

shows that those constraints should not be described as a content-bias embedded in the physiology of local neural structures.

As mentioned above, recent work from Dehaene and colleagues on the constraints involved in letter recognition in the VWFA displaces the content constraint view, and is, therefore, no longer in tension with the apparent preponderance of task-scale neural reuse. The alternative view focuses on facts about connectivity, such as the relationship in the ventral stream between lateral position and degree of foveal input, or the question of whether a site projects to language areas. Similar facts about the connectivity profile of the VWFA had been discussed in earlier work (Dehaene, 2009; Hannagan et al., 2015). However, in that earlier work, discussions of connectivity appear alongside claims about content bias in the VWFA. Facts about connectivity are framed as an explanation for why the VWFA appears where it does. This explanation of VWFA location appears to be offered as a *supplement* to the theory of content bias in the VWFA, rather than as a replacement for it. In the most recent work (Dehaene-Lambertz et al., 2018), the notion of content bias is simply left out. New longitudinal data allowed Deheane-Lambertz et al. to look back in time at the specific voxels in each subject that later came to be the site in which the VWFA emerged.[9] It turned out that, in pre-literate children, those voxels display far less stimulus preference than had previously been believed. In light of this new data, the 2018 paper suggests that the connectivity profile of the VWFA not only explains its location in cortex; it also generates the expected constraints on orthographic symbol use.

I'll now consider an objection that will likely have occurred to anyone familiar with research on object-selective cortex. Isn't the FFA a good example of a structure that has always been largely dedicated to one kind of content, and which, therefore, could have undergone selection for its capacity to represent faces? And if it did undergo selection for its capacity to represent faces, shouldn't we say that the representation of face-like content is both the primary function of the area, and the source of at least some of the developmental constraints it confronts in modern humans? Two lines of response are available. One is that the FFA may simply be an exception. One could argue that task-scale neural reuse characterizes the functional architecture of *most* of the brain, but not the FFA. In fact, this suggestion is compatible with what I've said so far. The central claim in this section has a conditional form: *if* a structure has long been involved in the implementation of task-scale neural reuse, then it is unlikely that the structure was tailored by natural

[9] If you want to study the site at which the VWFA will appear in the brains of children who are currently pre-literate, you have to guess where it will appear in the future. Individual variability imposes a relatively low ceiling on the accuracy of such guesses. The Dehaene-Lambertz et al. (2018) study is the first to overcome this methodological difficulty.

selection for the representation of some particular class of content. If the antecedent of the conditional goes unsatisfied in a particular case, the truth-value of the consequent is dialectically irrelevant. However, this response may not be the best one. The fact that the FFA might be an exception does nothing to show that an appeal to face-like content is the most appropriate way to articulate the nature of the developmental constraints on the capacities of the cortical site. In this connection, it is worth noting that, in order for past content to serve as causal constraint on the range of secondary functions a neural structure can acquire, the physiological properties underlying the content bias must be *canalized*. That is, the structure must end up acquiring those properties even in developmental environments that lack content-specific perceptual triggers. Without canalization in this sense, primary functions could not delimit the space of representational objects, as Dehaene puts it, because eventually, alternative cultural environments would emerge, and invite the development of alternative neural phenotypes. Is the FFA canalized in this sense? Until recently, this question had been impossible to answer. This changed in 2017, however, when Mike Arcaro and colleagues used welder's masks to raise three monkeys in a faceless environment. At 200 days after birth, which was the last time that imaging was done before exposing the monkeys to a normal social environment, the site corresponding to the FFA in those monkeys had not developed a preference for faces (Arcaro et al. 2017). This shows that, even in the case of the FFA, constraints on the development of cortical structures are not best articulated in terms of some pre-theoretially familiar class of representational content.

6 Three Consequences of the Clash

Here I will briefly draw out three conceptual consequences of the clash between timescales. The first consequence concerns the paradox of reading. Recall that the paradox of reading consisted of two explicit claims, and one implicit assumption. The first claim says that writing is too recent an invention for either writing or reading to have played a role in human genetic evolution. The second claim says that the word identification is localized to a particular cortical structure. The implicit assumption was that localized content functions are necessarily driven by the genetic evolution of the species, rather than by learning and development. In light of the clash between timescales, we can see that the assumption deserves to be rejected. Localization of content always depends on the task demands imposed by the developmental environment.

The second consequence of the clash concerns the character of ancient primary functions. The upshot of the previous section was that the kind of primary functions required by the content constraint view are not evolutionarily plausible. What then is the status of ancient primary functions more generally? This is a difficult question, but I think we can say this

much: if the goal is to characterize just one function that captures the historical role played by a given structure, we will have to generalize over the wide variety of task-scale neural functions supported by that structure. According to this suggestion, ancient primary functions do exist, but are more abstract than the content-constraint view requires. Once we generalize over all possible task-scale functions, there is little reason to think that the resulting conception of neural function will be accessible by means of folk-psychological reasoning. If such abstract functions can be represented accurately, it will be by means of a more rarified and theoretical form of representation, perhaps one that draws on the language of computation. Only such an abstract conception of function could bring unity to the otherwise heterogeneous list of context-bound functions that a given structure will subserve over evolutionary history. Alternatively, one might say that the list of context-bound functions is not subject to *any* unifying principle, regardless of the degree of abstraction we are willing to adopt. The best one can do is to produce open-ended lists of context-bound neural functions. Context-bound functions (whether oriented toward a particular task or not) are useful for many scientific purposes (Burnston, 2016), but they are too disparate to serve as a foundation for an ancient primary function. According to the context-bound list suggestion, nothing in nature satisfies the concept of ancient primary function.

Regardless of which view of ancient primary functions one prefers, the meaning of the claim that a neural structure has been subject to neural reuse on an evolutionary scale turns out to be far less transparent an idea than it had at first seemed. The need for a more abstract characterization of neural function threatens the coherence of evolutionary neural reuse, because, as discussed in Section 2, reuse demands a degree of conceptual distinctness between functions. If a cortical structure primarily performs an abstract function articulated in domain-neutral terms, such as, for example, gain modulation, then any apparently novel functional activity will count as an instantiation of the same function in a novel context, rather than as the realization of new function per se.

I suspect that the initially intuitive impression given by the idea of evolutionary neural reuse depends on the intuitive familiarity of the content functions that are mistakenly presumed to serve as the relata in the reuse relation. If reuse is imagined to be a transition between two content functions, both of which are accessible to folk-psychological reasoning, it will appear as though we already understand what is involved in a transition from primary to secondary functions (even if the observational consequences associated with the instantiation of either function are vague or indeterminate, and that, as a result, we cannot precisely specify the empirical content of transition events). However, once we take seriously the idea that ancient neural functions cannot be captured in terms of dedication to, or specialization in, any content-type that would be readily accessible from a folk-

psychological stance, intuitions about the boundaries between neural functions wither away. As they wither, so does the intuitive status of evolutionary neural reuse itself.

How far should we take this skeptical reasoning? Should we go as far as to declare that any suggestion of evolutionary neural reuse is conceptually bankrupt? Certainly not. Reuse applies to the structures that compose the human brain just as it applies to every other biological trait. As Darwin put it: "Thus, throughout nature almost every part of each living being has probably served, in a slightly modified condition, for diverse purposes, and has acted in the living machinery of many ancient and distinct specific forms (Darwin, 1877, p.284)." An immediate implication of Darwin's assertion is that neural reuse, in particular, has been common. We can accept that implication without presuming that we already know what the relata of the neural reuse relation are. Moreover, as noted in the initial discussion of content functions, there are many kinds of non-content functions to which the argument developed here does not apply.

The third consequence of the clash is a rather subtle, but also rather useful disambiguation of a prediction Michael Anderson makes about the relationship between the evolutionary age of a neural function, and the amount of cortical real estate it recruits. The ambiguous form of the prediction is this: in both evolutionary and developmental time, newer functions will demand more cortical real estate than older functions. It is valuable to figure out exactly what this prediction says, because it is one of the central principles that lends falsifiable empirical content to the neural reuse framework. If we insist on agnosticism about the nature of the relata in the neural reuse relation, while remaining cognizant of the diversity of kinds of neural function, the ambiguity in Anderson's prediction becomes easy to see. The prediction can be interpreted in strong and weak forms. The weaker interpretation treats the two timescales independently, and can be expressed like this:

> **Weak prediction.** It will typically be the case that, (i) for any given pair of functions characterized on a developmental timescale, F1 and F2, if F1 demands more cortical real estate than F2, then F1 will have developed later than F2, and (ii) for any given pair of functions characterized on an evolutionary timescale, F1 and F2, if F1 demands more cortical real estate than F2, F1 will have evolved later than F2.

The strong interpretation collapses the two timescales together. It can be expressed like this:

> **Strong prediction.** It will typically be the case that if function F1 demands more cortical real estate than F2, it will have appeared after F2 both in developmental and evolutionary time.

The crucial feature of the strong interpretation is that it appeals to the same pair of functions on both scales. It is a neuroscientific application of the late 19th century biologist Ernst Haeckel's memorable pronouncement that "ontogeny recapitulates phylogeny."

In light of the clash between timescales, only the weaker of these two claims is justified. The primary functions that get stabilized on an evolutionary scale will be content-neutral. At the task-relevant scale, many of the functions temporarily instantiated by any given structure will indeed involve the representation of a particular kind of content. Typically, therefore, the functions recognizable at an evolutionary scale will not be recognizable at a task-relevant scale. If so, content-oriented neural functions comprise a domain in which, contra Haeckel, ontogeny does not recapitulate phylogeny. The content of cognition is less tethered by the capacities of our ancestors than a casual consideration of neural reuse would suggest.

7 Constraint and Liberation

Thus far, I have argued against the idea that evolutionary constraints on human brain function can be articulated in terms of representational content. One might accept this conclusion, but nevertheless insist that evolutionary-scale neural reuse entails that cognitive function is constrained in other theoretically interesting ways. After all, there is no denying that we have inherited identifiable neural structures from our ancestors, and that the capacities of those neural structures make cognition possible. I'll conclude with a brief examination of this proposal.

To explore this idea, it will help to articulate what a "constraint" amounts to, in the domain of brain evolution. To say that the ancient functional profile of a neural structure constrains its modern homologue is to say that the range of capacities associated with the modern structure is narrower than it would have been, had the ancient functional profile been different. But different in what way? Many alternative ancient functional profiles would surely have led to an alternative set of contemporary capacities, but not necessarily to a narrower one. What kind of alternative ancient functional profile must we imagine, in order to make plausible the idea that, had that alternative been profile been the actual one, we would today enjoy an even broader suite of cognitive capacities? Precisely because task-scale reuse has been part of our species for a long time, it is hard to know how to answer this question. Given the ancient provenance of task-scale neural reuse, neural structures

have long been capable of realizing a diverse list of functions. Moreover, it is not at all clear that nature has imposed a theoretical upper limit on either the length or the diversity of that list. So neural reuse at the evolutionary scale has not clearly constrained us; or at least not in any way that we can confidently point to. The structures that compose our brains are constrained by their evolutionary history, but only in the non-committal sense in which every biological structure is “constrained" by its evolutionary history. Neural reuse does not entail some special, additional kind of constraint.

What about the opposite view? Is there any sense in which evolutionary neural reuse has helped to lift, or at least soften, some of the constraints on our mental life? Anderson (2014) predicts that the late-evolving capacities that are distinctive of human cognition require more extensive reuse of neural structures than older, less distinctively human capacities. Primary examples include the reuse of motor circuits for language (Pulvermüller, 2005) and numerical cognition (Penner-Wilger and Anderson, 2013). This suggests that, in comparison with other species, humans have an unusually amplified capacity to reuse neural structures for novel cognitive ends.

This idea is suggestive. In a poetic mood, one might even be tempted to say that neural reuse has been a source of human freedom. This claim carries more philosophical baggage than the corresponding claim about constraint, but its intended meaning is not difficult to work out. Its meaning is approximately the inverse of the claim about constraint. To say that neural reuse has been a source of freedom is to say that our species, in virtue of having acquired an unusually amplified capacity for task-scale neural reuse, is capable of realizing a broader set of neural functions now than we would have been able to realize, had that amplified capacity for task-scale neural reuse never been acquired. The counterfactual invoked by this claim is easier to evaluate than the one invoked by the claim about constraint, since, in this case, the counterfactual refers to a comparatively close possible world in which only one property is absent. Moreover, in order to evaluate this counterfactual, one does not need to know exactly what our species would have looked like, had task-scale reuse not emerged. It would suffice to show that the cognitive repertoire of our species would have been radically smaller without it. Let us assume that, at the level of the whole organism, the number of cognitive tasks that a human can accomplish is a function of the number of tasks that local neural structures can support. Assume also that each task recruits a network of local neural structures. If these two assumptions are correct, then the number of cognitive tasks that a human can possibly undertake will be a combinatoric function of the number of tasks each local structure can support. When viewed that way, task-scale neural reuse has exponentially increased the number of tasks we humans can undertake, and in that sense, has indeed been a source of human freedom.

References

Anderson, M. L. (2007). Massive redeployment, exaptation, and the functional integration of cognitive operations. Synthese, 159(3):329–345.

Anderson, M.L., Kinnison, J., and Pessoa, L. (2013) Describing functional diversity of brain regions and brain networks. Neuroimage 73, 50-58.

Anderson, M. L. (2014). After phrenology. MIT Press.

Anderson, M. L. and Finlay, B. L. (2014). Allocating structure to function: the strong links between neuroplasticity and natural selection. Frontiers in human neuroscience, 7:918.

Arcaro, M.J., Schade, P. F., Vincent, J. L., Ponce, C.R., and Livingstone, M. (2017) Seeing faces is necessary for face-domain formation. Nature Neuroscience 20, 1404–1412.

Barack, D. L. (2017). Cognitive recycling. The British Journal for the Philosophy of Science, 70(1):239–268.

Bergeron, V. (2010). Neural reuse and cognitive homology. Behavioral and Brain Sciences, 33(4):268–269.

Burnston, D. C. (2016). A contextualist approach to functional localization in the brain. Biology & Philosophy, 31(4):527–550.

Changizi, M. A. and Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. Proceedings of the Royal Society B: Biological Sciences, 272(1560):267–275.

Chapman, P. D., Bradley, S. P., Haught, E. J., Riggs, K. E., Haffar, M. M., Daly, K. C., and Dacks, A. M. (2017). Co-option of a motor-to-sensory histaminergic circuit correlates with insect flight biomechanics. Proceedings of the Royal Society B: Biological Sciences, 284(1859):20170339.

Coltheart, M. (2014). The neuronal recycling hypothesis for reading and the question of reading universals. Mind & Language, 29(3):255–269.

Darwin, C. (1877). On the various contrivances by which British and foreign orchids are fertilised by insects. John Murray.

Dehaene, S. (2008). Cerebral constraints in reading and arithmetic: Education as a "neuronal recycling" process. The educated brain: Essays in neuroeducation, pages 232–247.

Dehaene, S. (2009). Reading in the brain: The new science of how we read. Penguin.

Dehaene, S. (2013). Inside the letterbox: how literacy transforms the human brain. In Cerebrum: the Dana forum on brain science, volume 2013. Dana Foundation.

Dehaene, S. and Cohen, L. (2007). Cultural recycling of cortical maps. Neuron, 56(2):384–398.

Dehaene, S. and Dehaene-Lambertz, G. (2016). Is the brain prewired for letters? Nature neuroscience, 19(9):1192.

Dehaene-Lambertz G., Monzalvo K., Dehaene S. (2018) The emergence of the visual word form: Longitudinal evolution of category- specific ventral visual areas during reading acquisition. PLoS Biology 16(3): e2004103.

d'Errico, F. and Colagè, I. (2018). Cultural exaptation and cultural neural reuse: A mechanism for the emergence of modern culture and behavior. Biological Theory, pages 1–15.

Disotell, T. R., & Tosi, A. J. (2007). The monkey's perspective. Genome biology, 8(9), 226.

Gaillard, R., Naccache, L., Pinel, P., Clémenceau, S., Volle, E., Hasboun, D., Dupont, S., Baulac, M., Dehaene, S., Adam, C., et al. (2006). Direct intracranial, fmri, and lesion evidence for the causal role of left inferotemporal cortex in reading. Neuron, 50(2):191–204.

Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. Social neuroscience, 3(3-4):317–333.

Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., and Dehaene, S. (2015). Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. Trends in cognitive sciences, 19(7):374–382.

Haueis, P. (2018). Beyond cognitive myopia: a patchwork approach to the concept of neural function. Synthese, 195(12):5373–5402.

Iriki, A. and Taoka, M. (2012). Triadic (ecological, neural, cognitive) niche construction: a scenario of human brain evolution extrapolating tool use and language from the control of reaching actions. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1585):10–23.

Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. Proceedings of the National Academy of Sciences, 107(25):11163–11170.

Kim, J. S., Kanjlia, S., Merabet, L. B., and Bedny, M. (2017). Development of the visual word form area requires visual experience: Evidence from blind braille readers. Journal of Neuroscience, 37(47):11495–11504.

Luque, N. R., Naveros, F., Carrillo, R. R., Ros, E., and Arleo, A. (2019). Spike burst-pause dynamics of purkinje cells regulate sensorimotor adaptation. PLoS computational biology, 15(3):e1006298.

McCaffrey, J. B. (2015). The brain's heterogeneous functional landscape. Philosophy of Science, 82(5):1010–1022.

Naschev P., Kennard, C. & Husain, M. (2008) Functional role of the supplementary and pre-supplementary motor areas. Nature Reviews Neuroscience 9(11): 856–869.

Oyama, S. (2000). The ontogeny of information: Developmental systems and evolution. Duke university press.

Parkinson, C. and Wheatley, T. (2015). The repurposed social brain. Trends in Cognitive Sciences, 19(3):133–141.

Penner-Wilger, M., & Anderson, M. L. (2013). The relation between finger gnosis and mathematical ability: Why redeployment of neural circuits best explains the finding. Frontiers in Psychology, 4, 877.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. Nature Reviews Neuroscience, 6, 576–582.

Ramirez, J.-M., Tryba, A. K., and Pena, F. (2004). Pacemaker neurons and neuronal networks: an integrative view. Current opinion in neurobiology, 14(6):665–674.

Rathkopf, C. (2017). Neural information and the problem of objectivity. Biology & Philosophy, 32(3):321–336.

Rathkopf, C. (2013). Localization and intrinsic function. Philosophy of Science, 80(1):1–21.

Reich, L., Szwed, M., Cohen, Laurent, Amedi, A. (2011). A ventral stream reading center independent of reading experience. Current Biology 21: 363-368.

Thiebaut de Schotten, M., Cohen, L., Amemiya, E., Braga, L. W., and Dehaene, S. (2012). Learning to read improves the structure of the arcuate fasciculus. Cerebral Cortex. 24(4): 989-995.

Zerilli, J. (2019). Neural reuse and the modularity of mind: Where to next for modularity? Biological Theory, 14(1):1–20.