M. Chirimuuta

Dept. History & Philosophy of Science

University of Pittsburgh

mac289@pitt.edu

Extending, Changing, and Explaining the Brain

Abstract

This paper address concerns raised recently by Edouardo Datteri (2009) and Carl Craver (2010) about the use of brain-extending prosthetics in experimental neuroscience. Since the operation of the implant induces plastic changes in neural circuits, it is reasonable to worry that operational knowledge of the hybrid system will not be an accurate basis for generalisation when modelling the unextended brain. I argue, however, that Datteri's no-plasticity constraint unwittingly rules out numerous experimental paradigms in behavioural and systems neuroscience which also bring about changes in the brain. Furthermore, I propose that Datteri and Craver's arguments concerning the limitations of prosthetic modelling in basic neuroscience, as opposed to neuroengineering, rests on too narrow a view of the ways models in neuroscience should be evaluated. I distinguish *organisational validity* of models from *mechanisitic validity*. I argue that while prosthetic models may be deficient in the latter of these explanatory virtues because of neuroplasticity, they excel in the former since organisational validity tracks the extent to which a model captures coding principles that are invariant with plasticity. Changing the brain, I conclude, is one viable route towards explaining the brain.

1. *Introduction – Extending the Brain*

The science-technology relationship is of particular interest in brain research. Basic neuroscience yields hundreds of thousands of publications annually, exploiting an impressive range of techniques from genetic engineering to functional neuroimaging. Yet the discipline lacks an overarching theory of brain function to unify the vast quantity of data collected, and neuroscientists focussing on single levels of investigation (e.g. cellular, molecular, or systems), share little common ground. At the same time, certain findings in basic neuroscience have fostered practical applications, including neural technologies with significant therapeutic and commercial potential. Much neural technology aims simply to control the operation of neurons, especially in cases of psychiatric and neurological disease where function is pathological. Other technologies aim to *extend* neural function, for example by engaging parts of the cortex in the control of robotic limbs. These *Brain Computer Interface (BCI*[1]*)* systems are the focus of this paper. The techniques are made possible because of the brain's lifelong capacity for plasticity, the alteration of brain anatomy and physiology in response to trauma, demands of learning, or interaction with new objects in the environment. I discuss how such technologies can contribute to basic neuroscience as well as neuroengineering, asking if *changing* the brain can help in the project of *explaining* the brain. My answer will be roundly positive, in contrast to the views of two philosophers of neuroscience, Edoardo Datteri (2009) and Carl Craver (2010), who have challenged scientists' claims that BCI's can provide important insights into brain function which are not accessible via other techniques. Datteri and Craver's criticisms centre around the fact that hybrid BCI systems diverge from the natural systems basic neuroscience aims to model because of plasticity induced by prosthetic implants.

In support of my positive response I will argue for two distinct theses. The first, targeting Datteri, is that any concerns about experimentally-induced plasticity cannot be restricted to BCI preparations. Indeed, Datteri's concern about neuroplasticity over-generalises to numerous other paradigms in systems neuroscience because the same kinds of plastic phenomena occur both in BCI and non-bionic experiments (section 2.2). If this is so, and one takes the plasticity worry seriously, systems neuroscience is in trouble. The good news is that plasticity is only problematic on the assumption, accepted by Datteri, that the unique aim of neuroscientific research is to uncover the anatomically realized mechanism at play in the unextended, natural system (i.e. to build "how actually" models). Similarly, Craver's concern over the multiple realizability of functions carried out by natural and prosthetic systems rests on an assumption that basic neuroscientists must evaluate their models in terms of *mechanistic* and *phenomenal validity*, narrowly construed as the mapping of components, activities, inputs and outputs in the brain.

So my second main thesis (section 3.2) is that a more pluralist approach is needed in conceptualising the aims of neuroscientific research. Certain questions in systems neuroscience need not (and cannot)

[1] AKA *brain machine interfaces* (BMI). Craver (2010) calls these a kind of *prosthetic.* I often use Datteri's (2009) preferred terms, *bionic* and *hybrid.* For the purposes of this article, the terms prosthetic, bionic, or hybrid should be considered interchangeable in reference to models and experiments.

be answered with models that aim at a detailed specification of neuronal circuits. Instead, they concern what organisational principles hold across different circuits and are not tied to a particular realisation. I supplement Craver's phenomenal and mechanistic validity with the idea of *organisational validity*. When designing experiments to address questions of organisational principles – e.g. whether or not the motor cortex uses a distributed code – models are assessed for organisational validity, and plasticity in the experimental preparation is not a hindrance. My case studies illustrate that plasticity can actually be a help, so long as the preparation is not so grossly perturbed as to be performing its function in a fundamentally novel way.

Let us begin with a few words from Daniel Moran, professor of biomedical engineering at Washington University in St. Louis:

> "We'll drill a small hole in the skull, pop the bone out, drop the device in, replace the bone, sew up the scalp and you'll have what amounts to Bluetooth in your head that translates your thoughts into actions." (quoted in Lutz 2011)

The device in question is an epidural electrocorticography (EECog) implant, a recording device similar to an array of EEG electrodes but designed to rest on the cortex, inside the surface of the skull. It is one of a number BCI's in development for eventual clinical application in populations suffering from the most severe forms of paralysis due to malfunction of the motor nervous system. Users learn to adjust their patterns of brain activity so that the BCI provides real-time, voluntary control of a robotic limb, or moves a cursor on a computer screen. No residual motor skills are required, potentially restoring locomotive and communicative abilities to quadriplegic and "locked in" patients (Hochberg et al. 2012). To take just one example from Andrew Schwartz's laboratory at the University of Pittsburgh, monkeys trained with the BCI can use a robotic arm to reach to a marshmallow, grasp it in a pincer movement, and carry the food to the mouth (Velliste et al. 2008[2])

While BCI technology has received much attention for its great promise in rehabilitative medicine, it also has stood out as being of theoretical importance. It is tempting to assume that demonstrations of precise, engineered control over biological systems indicate that the system has been explained and understood. Perhaps this assumption tacitly fuels interest in neuroengineering. Dretske (1994) wrote, "if you can't make one, you don't know how it works". But this is not to say that if you can make one,

[2] Illustrative videos are available as supplementary materials at the *Nature* website, http://www.nature.com/nature/journal/v453/n7198/suppinfo/nature06996.html

then you *do* know how it works. Practical mastery may be a necessary, but certainly not sufficient, corollary of theoretical insight. This is a concern raised by Craver (2010). He claims that regarding the explanatory goals of basic neuroscience, prosthetic or bionic models implemented by BCI's do not have any advantages over standard ways of building models in neuroscience (841). Moreover, he argues, the bionic system does things differently from the natural system, so cannot constrain models of processing in the natural system (847). According to Craver, engineers and basic biologists find themselves in pursuit of different goals. Engineers' models aim at practical utility by any means, whereas biologists' models aim to mirror the workings of nature (850). One purpose of this paper is to show, contrary to Craver's claim, that BCI techniques *are* epistemically privileged with respect to certain kinds of questions about neural coding and organisation.

Datteri (2009) recommends even greater scepticism about the theoretical importance of experiments involving hybrid components, and asks what methodological constraints need to be imposed on such experiments in order that their findings can rightly contribute to basic neuroscience. Though curiously one of his constraints – that "one has to exclude that bionic implantations produce plastic changes in the biological components of the system" (305) – patently cannot be met by most known BCI technologies, because they depend on the capacity of neural tissue to adapt to the interface. This issue of the epistemic significance of neuroplasticity is really the crux of this paper, and my goal is to show how the methodologies and results of BCI experiments ought to be interpreted, such that plasticity cannot be said to compromise the theoretical significance of the research[3].

Before considering Datteri's and Craver's critical arguments in turn, it is necessary to say something about how exactly artificial interfaces extend and change the brain. I do not consider the extended *mind* in Andy Clark's sense, i.e. extending the mind beyond the bounds of the skull (Clark and Chalmers 1998, Clark 2004, 2008). Bionic devices may arguably do that, and one could think of the brain-prosthesis hybrid system as constituting an extended mind. However, the concern of this paper is with what happens to the *brain* following its interface with the artificial component. The brain is extended in

[3] Neuroplasticity is an umbrella term for countless types of phenomena. e.g. *synaptic plasticity* (changes in connection strength between neurons), *neurogenesis* (growth of new neurons in adulthood), *perceptual learning* (experience driven changes in sensory neurons' response properties), functional *reorganisation* following brain injury, so without further specification Datteri's target is somewhat amorphous. In what follows, the key phenomenon is the local reorganisation of motor cortex circuits (e.g. changes in neurons' excitability, tuning and connectivity) that correlates with improvements in performance while operating the robotic arm. This kind of plasticity is believed to be qualitatively similar to the sort that accompanies motor skill learning without prosthetics. It is important to emphasise that it is precisely these kinds of plastic effects that Datteri intends to exclude.

the sense that its repertoire of functions is expanded beyond the limits that are set by the facts of the brain's embodiment[4]. Crudely put, the brain's situation within the body, and its typical pattern of connections with sensory organs, the central nervous system, and muscular architecture, define a range of brain functions in relation to these relatively fixed "inputs" and "outputs"[5]. Adding a new kind of interface on the motor or sensory side allows for a new range of brain operations not possible within the unmodified bodily framework.

One might be concerned to distinguish neurotechnologies that merely provide a new interface for the brain, and those which change its inner workings enough to say it functions in a whole new way. For instance, if one puts an adaptor on an electric plug, the plug can now be used in different countries in virtue of this new interface, but its basic function remains the same[6]. So in principle, a new interface does not entail a new function. It pays, therefore, to consider the difference between the biological brain and artefacts like the plug in order to see why adding interfaces does, generally speaking, extend the range of functions of the brain. Importantly, the kind of input (an electric current) that your plug gets through the new adaptor is exactly the same as previously, which means that no components inside the plug need to be swapped or altered in order for the plug to operate with the input provided by the adaptor. With the BCI's currently available, on the other hand, the kind of input or output that these make available to the brain are different enough from the naturally occurring ones that some significant reorganisation has to take place within the brain in order for the interface to be usable. Fortunately, the brain has an inherent capacity for reorganising itself which means that direct experimental intervention on areas of the brain connected to the interface are not required in order for the equivalent of component swapping to occur. Thus the BCI should not be thought of as merely an additional interface (e.g. a plug adaptor), but also as something that modifies the inner workings of the brain and thus adds new functions (e.g. from a simple electric plug to a voltage transformer) [7].

[4] This clarification is needed because there is a sense in which any skill learning extends the brain beyond its previous repertoire of functions – e.g. learning to type, play the violin. I do not assume that there is a difference in kind between the kinds of brain plasticity and extension of function required for skill learning, and those observed following BCI use. It is just that the latter will not be observed in the absence of specific technological interventions because they rely on new kinds of brain-implant-body connections offered by the technology.

[5] Scare quotes because I do not aim to reinforce the simplistic picture of the brain as sandwiched between sensory inputs and motor outputs (see Hurley 1998).

[6] Thanks to an anonymous reviewer for raising this concern and for suggesting the electric plug metaphor.

[7] To pursue the electric plug metaphor, imagine an electrical motor built in the UK and designed to operate on a 240 V supply. Using a standard plug adaptor, the device is switched on in the USA and because it now only has a 110 V supply it doesn't operate at full speed. But this device has an inherent capacity to modify internal components in response to the demands of the new electrical input, and in time begins to run as it did in the UK. It behaves as if it has grown an internal step-up transformer. This is what the brain is like as it adapts to the BCI.

This kind of change is most easily illustrated with sensory substitution technologies which interface with the "input end" of the brain, and rely on the brain's ability to adapt to a different format of sensory information. The cochlear implant which stimulates the auditory nerve is the most successful BCI to date. Even though it is designed to mimic the activity of the cochlear, the kind of input it provides has vastly fewer frequency channels than would occur naturally. For this reason, the auditory cortex must undergo a process of adaptation to make best use of the artificial input and recover intelligible speech. For the congenitally deaf, cochlear implantation is most successful if introduced before two years of age, when the brain is most plastic, meaning that entire regions of the cortex can be co-opted for new purposes that might otherwise be given over to non-auditory modalities (Harrison et al. 2005). Tactile-visual sensory substitution (TVSS) has been much discussed as a potential means of restoring sight to the blind by the re-routing of optical information through the touch receptors of the skin (Bach-y-Rita 1972, Lenay et al. 2003). Extensive training is required for the use of TVSS, and neuroplasticity is recognised to underlie this process as the brain reorganises itself in order to utilize the new kinds of inputs (Ptito et al. 2005). In this sense TVSS extends the brain – it prompts the brain to reinforce and forge new pathways from peripheral somatosensory nerves to the visual cortex, therefore expanding its repertoire of functions.

At the "output end", devices which are designed to control artificial limbs usually interface with the primary motor cortex (M1). Since activation in this brain area normally brings about movement in a healthy person, it might be thought that this case is more analogous to the simple plug example, because it would seem that the brain just needs to do what it does normally in order to control a robotic rather than a real arm.[8] But as Andrew Schwartz describes in an interview with a science journalist, this is not so:

> "Although there is an area of the cortex generally associated with arm motion, the exact placement of the electrodes is not crucial, Schwartz explained. 'You don't have to be exactly right because the brain is highly plastic', he said, referring to the fact that the brain will rearrange its structure to get things done. …. 'Our algorithm is not exactly what is going on in the brain,' Schwartz said. But the monkey's brain actually adapts its neural signal to be closer to the algorithm." (Schirber 2005)

The point is that even if one places the BCI electrode exactly on the arm area of M1, adaptation will

[8] Note, however, that it need not be movement in an artificial body part that is generated, since many BCI experiments just require subjects to control the movement of a cursor on a computer monitor; and also, it has been shown that parts of the brain other than motor cortex can be co-opted for this purpose (Leudthardt et al. 2011).

still be necessary because the algorithm used to recover a meaningful motor command from the neural activity recorded itself introduces biases that the brain must compensate for (Koyama et al 2009). So since the brain has to adapt anyway to the specific way that the prosthetic system responds to motor commands, the researchers do not bother trying to place the electrode specifically on the arm area. They let plasticity do the work[9], and the prosthetic system effectively co-opts neural circuits for tasks that go beyond their former functional range[10].

The effects of these functional extensions are not instantaneous. A certain period of training is required before performance in the BCI motor control task is satisfactory in terms both of speed and accuracy (see Taylor et al., 2002; Musallam et al., 2004; Schwartz, 2007; Ganguly and Carmena, 2009). This is related to the time needed for neuroplastic changes to occur within the brain. As Legenstein and colleagues (2010:8400) write,

> "Monkeys using BCIs to control cursors or robotic arms improve with practice, […] indicating that learning-related changes are funneling through the set of neurons being recorded."

Studies have also measured the time course and extent of BCI induced changes in the activity profiles of individual neurons and populations and related these to behavioural findings (e.g. Carmena et al. 2003; Jarosiewicz et al. 2008). A crucial point – one that appears to have been overlooked in Datteri's writing on this subject – is that the BCI induced plasticity is not qualitatively different from learning related plasticity occurring in the absence of technological intervention. It is well known that strength of synaptic connections, number of long range connections, and response properties of individual neurons are all rapidly modified with perceptual and behavioural experience (see Shaw and MacEachern 2001 and Pinaud et al. 2006 for overviews). And as Sanes and Donoghue (2000: 393)

[9] It might be suggested that a prosthetic that used both accurate electrode placement *and* a more naturalistic decoding algorithm would have no need to rely on cortical plasticity. In a follow up to this paper (Author, in preparation), I explain the practical and theoretical limitations on making decoding models maximally realistic in this way.

[10] One may also object to my claim that BCI's functionally extend the motor cortex by suggesting the alternative hypothesis that the co-opting of circuits for the new tasks is just normal re-use (see Anderson (2010) on the neural re-use hypothesis; and thanks to an anonymous reviewer for raising this suggestion). As it happens, there are grounds for thinking that some phenomena commonly attributed to plasticity may actually be instances of re-use – e.g. that M1 has been described by different labs as encoding abstract direction of movement or controlling muscle activity, depending on experiments performed in those labs (Meel Velliste, personal communication). According to the simple plasticity account, one or both of these functions is not naturally performed by M1, and it must learn to do it; but it *could* be that M1 is able to perform and switch between both of these functions even under non-experimental conditions. Importantly, the reuse hypothesis does not predict there will be structural changes in neural circuits called on to perform different functions. However, what is clear from the literature on BCI's, and normal motor skill learning, is that such changes are also taking place, e.g. in the form of alteration of motor cortical neurons' directional tuning preferences and domain of control (see Sanes and Donoghue 2000 for review), and so such effects are universally considered as instances of plasticity. It is these phenomena that I focus on.

write, “MI has a plastic functional organization in adult mammals. This property ostensibly results from a broad connectional organization, and the capacity for activity-driven synaptic strength changes.” They go on to link this capacity to behavioural data on motor skill learning, noting the “remarkable flexibility in motor behavior” (396) of primates and other mammals that this plasticity makes possible. In what follows, the kind of plasticity in question can be characterized, roughly, as the neural correlate of motor learning.

Now this phenomenon may not appear to be of great import – it is not news that people and animals learn new skills, and unless you are a substance dualist you would expect something to be occurring in the brain to make this possible. Yet, both Datteri and Craver have concluded that plasticity must be an obstacle to the project of explaining the brain in BCI research. In what follows, I will show why such concerns over plasticity are misplaced. One purpose of this paper is to describe in greater detail the role of plasticity in neurotechnology and the prevalence of plastic phenomena in basic neuroscience. Another is to examine scientists' own claims for the theoretical significance of progress in neuroengineering, arguing in response to Datteri and Craver that the methodological norms suggested by the scientists are reasonable. Section 2 examines Datteri's no-plasticity constraint, arguing in section 2.2 that it overgeneralises to many experimental protocols in systems neuroscience. Section 3 considers Craver's account of the differences between the goals of basic neuroscience and neuroengineering. In section 3.2 I argue that a more pluralist conception of the aims of research can encompass the uses of brain computer interfaces in basic neuroscience. Section 4 discusses some further issues which place this topic of BCI's and neuroplasticity in the context of other debates in the philosophy of neuroscience, and philosophy of science more generally.

2. *Changing the Brain in Experimental Neuroscience*

In this section I examine Datteri's cautionary observations regarding bionic preparations which induce plastic changes. Datteri's assertion of the need for a “regulative methodological framework” (301) for the use of BCI's and related technologies when modelling biological systems is stronger than Craver's central point, that prosthetic models are not epistemically privileged. Both philosophers can be understood as a reacting against certain expectations raised by scientists engaged in BCI research. In fact, Datteri quotes two papers from Miguel Nicolelis' laboratory at Duke which herald the arrival of the BCI as a core technique in computational and behavioural neurophysiology:

> ''the general strategy... of using brain-derived signals to control external devices may provide a unique new tool for investigating information processing within particular brain regions'' (Chapin et al., 1999, p. 669).
>
> ''[Brain-computer interfaces] can become the core of a new experimental approach with which to investigate the operation of neural systems in behaving animals'' (Nicolelis, 2003, p. 417).

Nicolelis (2003) calls this new experimental technique "real-time neurophysiology" because the BCI preparation involves high resolution recording of neural activity in primary motor cortex and simultaneous decoding of the activity for immediate use in control of a robot arm.

2.1 *Datteri's Plasticity Worry*

The technique is obviously promising, but Datteri's concern, simply put, is that the data one obtains through real-time neurophysiology will contain artefacts due to the presence of the implant and will not shed light on ordinary mechanisms for motor control. If plasticity is taken to be one such artefact, then there is certainly a problem with this experimental method because, as we have seen, plastic changes are a pervasive feature of BCI research and are actually required for the correct functioning of the technology. However, Datteri's no plasticity caveat – that before drawing conclusions from BCI research for basic neuroscience, "one has to exclude that bionic implantations produce plastic changes" (2009:305) – is oddly out of joint with the actual business of neuroscience; not least because the plasticity occurring in response to the BCI is not qualitatively different from plastic changes occurring in non-bionic experiments. In some circumstances philosophers of science may raise legitimate concerns about methodological conventions that have been applied unquestioningly by scientists[11]. I contend that this is not one such case. For as will become clear in section 3, Datteri assumes an evaluative framework that is too limited to be applicable to systems neuroscience research. But first I will discuss his claims in more detail.

The part of Datteri's paper that I will focus on is his discussion of experiments which involve the replacement of a piece of neuronal circuitry with an artificial component. Here, the resulting hybrid system (called an "ArB" – "Artificial replaces Biological" system) can be used to test a hypothesis

[11] I will return to this issue in Section 4 below.

about the properties of the substituted biological part. Drawing on Craver's (2007) account of mechanistic explanation in neuroscience, Datteri envisages the goal of BCI research as getting from "how-plausibly" simulations of the biological system to "how-actually" models (305).

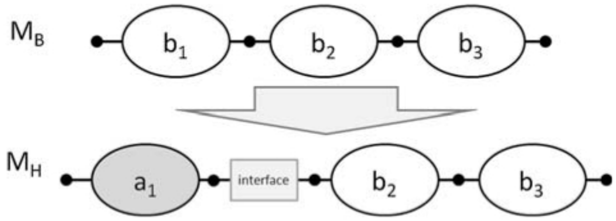


Figure 1 (Datteri 2009 Fig 4. Permission required)

M_B is the mechanism description of a biological system and b_n are the biological components linked together at nodal points. M_H is the mechanism description of a hybrid system and a_1 is artificial component replacing b_1, linked to the other biological components by the interface.

Figure 1 is a schematic which Datteri uses to describe the relationship between a model ("mechanism description") of the untampered biological system (M_B) and a model of the hybrid system(M_H) in which the first component is replaced by an artificial substitute, interfacing with neurons through a BCI. Datteri (2009:310) explains how a hybrid ArB system may be used to study the relationship between the model of the biological system, and the system itself . In particular, it serves to test the hypothesis that a component b_1 behaves as M_B describes. The concrete example given is Zelenin and colleagues' (2000) model of the reticulo-spinal pathway in the lamprey, a system which controls stabilisation of the body during swimming. In this case, the hypothesis tested by the bionic system is concerns how each recticular neuron controls rolling movements (Datteri 2009: 311-2). The lamprey is fixed on a motorised platform and the usual pathway from the recticular neurons to the spinal neurons controlling movement is severed. Activity in the recticular neurons is monitored with a brain computer interface and the neural firing is decoded to predict the 'intended' roll movement generated by the neurons. This signal then governs the tilt of the platform. The key intervention in the experiment is to initiate off-balance motion in the motorised platform (i.e. a movement not governed by the recticular neurons) and observe how recticular activity responds to stabilise the animal. The fact most important to our discussion is that this experimental preparation is schematically equivalent to the motor BCI preparation in primates: the firing of recticular neurons is equivalent to M1 activity, and stabilising movement of the platform is equivalent to movement of the robotic arm. In both cases, the usual pathway from neuronal activity to bodily movement is blocked and the animal's 'intended' movement is decoded by an external computer which receives neuronal activity via a BCI and then controls movement in an artificial motorised device.

Datteri argues that certain conditions must be in place in order that the hybrid system can truly be said to inform scientists about the biological one, and is rather forceful about the implications if they are not met. If these conditions cannot be satisfied, Datteri writes, "one may reasonably doubt that the analysis of hybrid system performances can play a significant role in the scientific modelling of the target biological system, thus suggesting the need for adding crucial qualifications to Chapin's and Nicolelis's claims" (315). The conditions are:

> (ArB1) the brain-machine interface included in the system does not introduce uncontrolled perturbations which interferes [sic] unpredictably with normal mechanism working. (310)
>
> (ArB2) H and B are identical except for the replaced component. (311)
>
> (ArB3) artificial component a_1 is governed by the regularity which, according to M_B, governs the behaviour of the corresponding (replaced) biological component b_1. (311)

I shall say little about the first and third of these until section 3.2 below. I focus on ArB2 because it strictly and explicitly rules out the theoretical value of any BCI experiments that involve neuroplasticity. Concerning ArB2, Datteri writes that in order to, "draw ..[a] theoretical conclusion on the basis of behavioural similarities between the bionic and the biological system… [o]ne needs also to assume that the biological, non replaced part of [e.g.] the lamprey has undergone no changes as effect of the bionic implantation" (312). But all of the applications of BCI's to date in humans and other mammals, both at the "input" stage (sensory substitution) and at the "output" stage (motor control), have relied on some degree of reorganisation of the neural circuits interfacing with the device. This constraint rules out a vast swathe of BCI research as not informative in the modelling of actual biological systems for sight, hearing or reaching. To make this clear I will now address a pair of issues about how one could or should apply Datteri's condition to experiments involving mammalian motor cortex.

First, the sense of "identical" in ArB2 is left unspecified. It could have a weaker sense of functional equivalence, or a stronger sense of being effectively indistinguishable in anatomy and physiology. Functional equivalence is often conserved across plastic changes, so a brain area modified after undergoing a BCI experiment could still be identical with its former self in this sense. For example the motor cortex of a laboratory monkey may be functionally equivalent in reaching tasks before and after BCI training, but have undergone significant reorganisation neural circuits. Yet it is doubtful that this weaker sense could apply because if so this would not, by itself, exclude preparations where plasticity occurs, and Datteri explicitly highlights as problematic[12]. So I take it that Datteri must have the stronger sense in mind.

Second, if the artificial component (a_1) exactly mimics the input-output operations of the biological one (b_1), then there is presumably no need for rest of brain to adapt itself to meet it. One may wonder if Datteri's constraints specifically target these kinds of systems, for this is how Datteri describes the lamprey case. By contrast, in motor cortex BCI preparations inputs and outputs through the interface differ substantially from naturally occurring ones. But given that there is no schematic difference between the two kinds of systems (fig. 1 characterises both of them equally well), it is by no means uncharitable to read him as targeting both. Moreover, Datteri explicitly presents motor cortex BCI's as examples of systems that contravene ArB2, referencing work by Hochberg and colleagues (2006), and from the Nicolelis group. To put this in context, it is worth noting that the examples of good input-output matching, like the lamprey, are a rarity in BCI research. If Datteri's constraint targets only those systems, then it would mean that his regulative framework has nothing to say about the majority of BCI experiments; yet it is presented as a general framework for evaluating bionic research. I note also that the scientists' claims for the importance of BCI tools in basic neuroscience, that Datteri highlights as problematic, are both from motor cortex BCI papers. I conclude that ArB2 is intended to apply to all motor BCI systems, regardless of differences in the degree to which a_1 mimics the inputs-outputs of b_1[13].

[12] "Second, as far as ArB2 is concerned, many studies show that bionic implantation is likely to produce long-term changes in the biological system. It has been widely demonstrated … that the implantation of a bionic interface and the connection with external devices typically produces plastic changes in parts of the biological system, such as long-term changes of neural connectivity. Other plastic changes affect the activity of neurons." (313)

[13] One could of course argue that Datteri ought to have treated the lamprey and the motor cortex cases differently because of the difference in degree of input-output matching, instead of lumping them together. In effect, that is to concede my point that Datteri's framework is inappropriate for most of BCI research. It does seem, however, that Datteri underestimates the prevalence of BCI's showing poor input-output matching, and the importance of plasticity for the working of most BCI's. He writes that in the case of M1 interfaces, ArB2 is likely to be contravened by *undetected* changes just because the initial

So as it stands, the no-plasticity constraint rules out all existing motor BCI preparations as useful for the modelling of motor systems in the brain because of neuroplasticity. This appears implausibly restrictive and uncharitable to the actual activities of BCI researchers, but that is not reason alone to reject the constraint. In the next subsection I will argue that ArB2 overgeneralises in a way that makes its application unacceptable in systems neuroscience, and in section 3.2 below I will discuss how important results for basic neuroscience come out of BCI experiments because of plastic effects.

2.2 *Neuroplasticity in Non-bionic Experiments*

To summarise the issue at hand, experiments involving BCI's for motor control have been presented by some of the scientists involved as an exciting new way to understand neural processing for motor control, whereas Datteri urges caution over such claims because the neuroplasticity occurring in these experiments means they cannot meet a condition specifying conformity between the unextended and the hybrid systems. This raises the question of whether more credence should be given to the neuroscientist's enthusiasm or the philosopher's caution. As philosophers of science should we operate a principle of charity when examining scientific methodology? If our analysis rules out an entire programme of research (i.e. the use of motor cortex BCI's in basic neuroscience), is that reason enough to reject the constraints imposed by our analysis? If the issue were just confined to the minority of experiments involving BCI's, that might not be sufficient grounds alone for challenging the no plasticity constraint (ArB2). Yet we can see that the problem for Datteri is even more widespread since his negative conclusions generalise to a substantial proportion of research done in mainstream systems neuroscience, not involving BCI's but inducing plasticity nevertheless.

Systems neuroscience is the field which tries to understand how the interrelations of large numbers of neurons bring about perceptions, motor responses, emotions, cognition, etc.. This research involves a combination of methods borrowed from psychology (e.g. visual psychophysics, working memory

state of the biological system is less well characterised and so "plastic changes may be hard to detect and predict due to the lack of adequate theoretical models" (315). But from what is known already about the way that such techniques extend brain function, there is no question of any researchers being unaware of plastic changes they induce in motor cortex! Datteri neglects the importance of plasticity to the actual working of the BCI. To reiterate, functioning prosthetic implants are possible *because* the brain adapts to them.

probes) which precisely measure behavioural effects of brain activity, and physiological methods (e.g. fMRI, electrophysiology) which record neural activity more directly. The crucial point is that plastic effects are not rare occurrences only observed in BCI laboratories, but they are almost omnipresent in systems neuroscience research. Plasticity is the neural accompaniment to any kind of experiment involving a behavioural task which is subject to increased performance with training during laboratory sessions. Memory and skill learning need not be the explicit targets of investigation, but almost any task can elicit improvement in a sensory, cognitive, or motor capacity with a small number of practice trials. To the extent that any experiment in systems neuroscience involves the subject learning a specific task in the lab, there will be subtle, but real, changes happening in the brain.

For example, most behavioural tests in visual neuroscience do not involve naturalistic seeing, but the measurement of discrimination or detection thresholds for novel artificial stimuli. Thresholds typically go down with practice until training is complete. Perceptual learning is correlated with changes in the visual cortex such as differences in neurons' receptive field size and organisation, and can be observed with a wide range of experimental paradigms (see e.g. de Weerd et al. 2006, Kourtzi 2010, Sagi 2011). Note that amount of plasticity accompanying perceptual learning has probably been underestimated in the past (Chirimuuta and Gold 2009), though it has always been recognised to some degree. However, in a systems neuroscience experiment that does not focus on learning and plasticity specifically, neural activity is typically recorded only after training. This means that the data are collected from a brain that is not the same as it was before its introduction to the laboratory. In a clear sense, it is an experimentally modified brain, analogous to the brain that has been modified due to the introduction of a bionic implant.

On the strength of the fact that both bionic and non-bionic preparations can change during experimental procedures, then if the no-plasticity constraint applies to one it should also apply to the other. This argument needs tightening, for of course there are differences between the hybrid experimental setup and the non-bionic one. In order to make the comparison, we must recognise that a key assumption behind Datteri's account is that the aim of the BCI experiment is to provide either negative or supporting evidence for a particular mechanistic model of the biological system (Datteri 2009:305, cf. Craver 2010:843). The model aims to specify the relevant features of the biological mechanism (e.g. in the number, arrangement and activities of its parts) so if the experimental preparation used to test the model diverges from the biological system in more ways than just an obvious replaced part or implant, then it can no longer aid in the model's reconstruction of that system. This mirroring of model and neural mechanism is what Craver (2010) calls "mechanism validity" (see section 3 below). Just as in the BCI case, a standard neuroscience experiment which causes plastic modifications to the target system will not be a reliable means to construct a model of the mechanism as it existed before the changes. So if representing in this sense is the only goal of research, plasticity is as much of a problem for non-bionic as it is for bionic methods.

The use of a bionic implant is just one way of getting round the problem that one cannot directly examine a neurobiological mechanism, *in toto*, as it operates in the natural setting. Instead, components of the model have to be localised and isolated (following the heuristics of decomposition and localisation described by Bechtel and Richardson 1993) and prepared in the laboratory for testing. One approach to testing the model's predictions regarding a particular component is to produce a hybrid ArB system with an artificial replacement of that component. A more straightforward way is to collect data from the isolated component itself, e.g. through electrophysiological recording or neuroimaging, and compare these with the model prediction. This is the approach taken in most of systems neuroscience research.

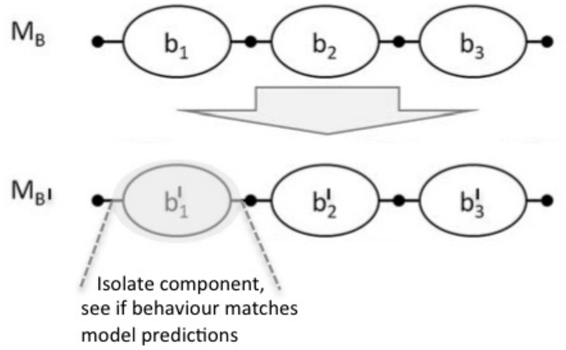


Figure 2 –Model of biological system and its experimental preparation

M_B is the mechanism description of a biological system. It cannot be tested directly. $M_B{}^I$ is the mechanism description of the experimental preparation of the biological system. Components $b_n{}^I$ can be tested in isolation from the rest of the system.

Figure 2 shows the relation between the two models in a way comparable to the relation between MB and the hybrid model depicted in figure 1. The model or mechanism description of the original biological system (MB) should be labelled differently from the model of the laboratory preparation

(MB^I). The fact that the biological system must be prepared in some way before it can be tested opens up a space between the original system and the one that is experimented on, such that it cannot be guaranteed that the two are identical. In this context, one obtains the following no-plasticity constraint which is equivalent to ArB2*:*

($B^I rB$) B^I and B are identical with respect to the component undergoing testing.[14]

Yet many standard experimental preparations cannot satisfy this condition. For example, if component b_1^I is physically isolated from the rest of the brain (*in vitro* slice preparation) for the purposes of intracellular recording, it cannot be guaranteed that its behaviour will not be different from the original b_1. Similarly, the function of contrast discrimination is effectively isolated in vision experiments, by presenting simple stimuli such as black and white gratings which only figure contrast information, without colour, complex 3D structure, etc. (see e.g. Legge and Foley 1980, Holmes and Meese 2004) and by training the subject to perform with an optimal degree of accuracy at this specific task. As David et al. (2004) report, the neurons responsible for contrast discrimination then begin to behave in a way which is subtly different from their operation under natural conditions.

Schematised in this way, the instances of non-bionic and bionic induced change are truly comparable. The only difference is that in the bionic case the concern is with plasticity in the other components of the system that have not been replaced (b_2 and b_3), whereas in the non-bionic cases described, plasticity is induced in the target component (b_1), but may consequently affect other components in an *in vivo* preparation. So if Datteri is correct about the need for a stringent no-plasticity constraint in BCI research, he has inadvertently hit upon a major flaw in systems neuroscience. That is a big *if*. In fact, plasticity is only a problem on the assumption that the unique aim of the experiment is to get a detailed, quasi-anatomical, account of a static, unchanging neural mechanism. My argument in the next section is that the aims of experiment must be construed much more broadly than this. By examining in more detail what the neuroscientists claim to have learnt from BCI research about the workings of the motor cortex, we will see that plasticity is not an obstacle to these alternative goals.

[14] Following p.11 above, the sense of "identity" here is that of having anatomical components and physiological properties that are effectively indistinguishable for the scientists comparing the systems. A plastically modified system will not be identical, in this sense, to the original one.

3. *More Ways to Explain the Brain*

This section will present an alternative way of conceptualising the aims of neuroscientific research. I argue that there are more ways to explain the brain than have been assumed by Datteri and Craver in their criticism of BCI methods. Conceptions of scientific aims impact on how the research will be evaluated, what kinds of explanation are derived from the findings, and they ultimately decide whether plasticity is to be considered an epistemic problem. Carl Craver (2010) in fact lists five evaluative dimensions for models, simulations and prosthetics in neuroscience: *completeness, verification, phenomenal validity, mechanistic validity* and *affordance validity*. I will argue that even though this list covers much ground, it lacks the conceptual resources to accommodate the ways that BCI research can contribute to basic neuroscience. Section 3.2 presents scientists' claims for their contribution to basic neuroscience. These do not amount to the specification of an actual circuit or mechanisms e.g. for motor control. Instead, they arrive at more abstract principles, that can be applied across mechanisms that are changing plastically. I conclude that prosthetic models should be evaluated according to a new dimension, *organisational validity*, which assesses the extent to which a model or explanation encapsulates invariant organisational principles, regardless of circuit reorganisation.

3.1 *Varieties of Validity*

Craver's "Prosthetic Models" paper (2010) focuses on the epistemic value of models involving BCI's asking, "What if anything does the effort to build a prosthesis contribute to the search for neural mechanisms over and above the more familiar effort to build models and simulations?" (840). Importantly, he conceives evidence to be any finding that constrains the space of possible mechanisms for a phenomenon (843), so that the ultimate aim of research is to narrow the space of possibilities to one. He contrasts BCI experiments involving electronic circuitry designed to functionally augment or replace samples of neural tissue with those experiments which use electronics just passively to record the activity of neural ensembles[15]. Craver is particularly concerned with the three types of validity, thought of as "fit […] between a model and the world" (843). These are phenomenal, mechanistic and

[15] For simplicity of exposition, and consistency with the rest of the paper, I focus on Craver's example of the BCI for movement control, rather than the alternative case study of Berger's prosthetic hippocampus. The conclusions he draws are not different for the two examples.

affordance validity[16].

Phenomenal validity is the extent to which the model's "input-output function is relevantly similar to the input-output function of the target" (842), while a model is *mechanistically valid* if "the parts, activities, and organizational features represented in the model are relevantly similar to the parts, activities, and organizational features in the target"[17] (842). So both phenomenal and mechanistic validity require a correspondence between the relevant features of the model and the target biological mechanism. *Affordance validity,* on the other hand, is non-representational. It is simply "the extent that the behavior of the simulation could replace the target in the context of a higher-level mechanism" (842). That is, it must function in a satisfactory way.

The nub of Craver's discussion is that while prosthetic models excel with respect to affordance validity, this is no guarantee of phenomenal or mechanistic validity:

> "Consider mechanistic validity first. Prosthetic models at their most biologically realistic are engineered simulations. As such, they inherit the epistemic problem of multiple realizability[18]. A prosthetic model might be affordance valid and phenomenally valid yet mechanistically invalid. Prosthetic runners legs do not work like typical biological legs. Heart and lung machines do not work like hearts and lungs. If so, then building a functional prosthesis that simulates a mechanistic model is insufficient to demonstrate that the model is mechanistically valid." (845)

Concerning phenomenal validity, Craver explains that the key difference between the biological

[16] The dimensions of completeness and verification describe how exhaustively and faithfully the model or simulation reproduces features of the biological mechanism. As Craver writes "All models and simulations of mechanisms omit details to emphasize certain key features of a target mechanism over others. Models are useful in part because they commit such sins of omission" (842). I will return to this point in section 4 below, and in this section concentrate on the three kinds of validity.

[17] Given that the topic is systems neuroscience, rather than cellular or molecular neuroscience which study sub-neuronal mechanisms, I understand the key "parts" here to be neurons, so that for a model of a brain circuit to be mechanistically valid it must be quite anatomically accurate, featuring the same number and type of neurons as in the actual mechanism.

[18] One wonders if Craver is saying that if multiple realizability were to occur in a non-bionic experiment, this would cause the same epistemic problem. In fact one cannot assume that mechanisms in systems neuroscience are not multiply-realized across individuals and across the lifespan. No two brains are identical, and circuits controlling perceptions and actions are sculpted and personalized by genetics and experience. It seems that the problem of failing to achieve mechanistic and phenomenal validity generalizes to non-bionic systems neuroscience, on Craver's analysis. This point is comparable to the one made above (section 2.2) that Datteri's no-plasticity constraint must apply to non-bionic experiments in systems neuroscience, if it is to apply to bionic ones. However a more charitable reading of Craver takes up the point that the *range* of inputs and outputs used by nature is much narrower than that use by engineers ("The space of functional inputs and outputs is larger than the space of functional inputs and outputs that development and evolution have *thus far* had occasion to exploit." p.847). Basic neuroscience, in its quest for phenomenal validity, can be said to be targeting this subspace of the expanse of possible inputs and outputs. Likewise, systems neuroscientists could be said to be working towards a description of the small range of mechanisms employed by different people for a specific function.

mechanism and the prosthetic model is in the pattern of inputs and outputs used. He notes that no existing BCI for motor control uses just those neurons that the brain uses to move the right arm. Furthermore, neuroplasticity means that the space of possible inputs and outputs is not tightly bounded (846, cf. 848). In both instances, the fact that the function which the prosthesis replicates is multiply realizable (due to plasticity) suggests to Craver that the epistemic value of prosthetic modelling is limited. Given that a vast range of internal mechanisms and input-output patterns can realize the same function, building the hybrid system cannot tighten the net around the actual mechanisms used in nature.

Craver makes a number of further points, centring on the contrast between basic neuroscience (i.e. "explanatory knowledge of how the brain works", 849) and neural engineering (i.e. "maker's knowledge of how to prevent disease, repair damage, and recover function." 849). Basic neuroscience, in its quest for explanatory knowledge is associated with obtaining models that are both phenomenally and mechanistically valid, while neural engineering settles for affordance validity and the maker's knowledge of "how the brain might be made to work for us" (840). In effect, BCI research does not help neuroscientists explain the brain. So even if Craver's first formulation of his thesis (as stated in the abstract), is quite weak – that prosthetic models provide a sufficient test for affordance validity and are, "[i]n other respects […] epistemically on par with non-prosthetic models" (840), one conclusion that he arrives at by the end of the paper is stronger:

> "I argue that affordance valid models need not be mechanistically or phenomenally valid. This is a blessing for engineers, and a mild epistemic curse for basic researchers." (850)

In other words, the failure of affordance validity to correlate with mechanistic and phenomenal validity means that a well running prosthetic model will probably shed no light on the workings of nature. As with Datteri's, Craver's stronger conclusion rests on a clear assumption about the goal of basic neuroscience, i.e. that the research should try to reveal actual mechanisms or circuits for motor control, rather than organisational principles that make a variety of different mechanisms or circuits function effectively. In section 3.2 below I argue that despite multiple realizability, BCI research can in fact contribute to the explanatory projects of neuroscience.

3.2 *Principles, Mechanisms and Explanatory Knowledge*

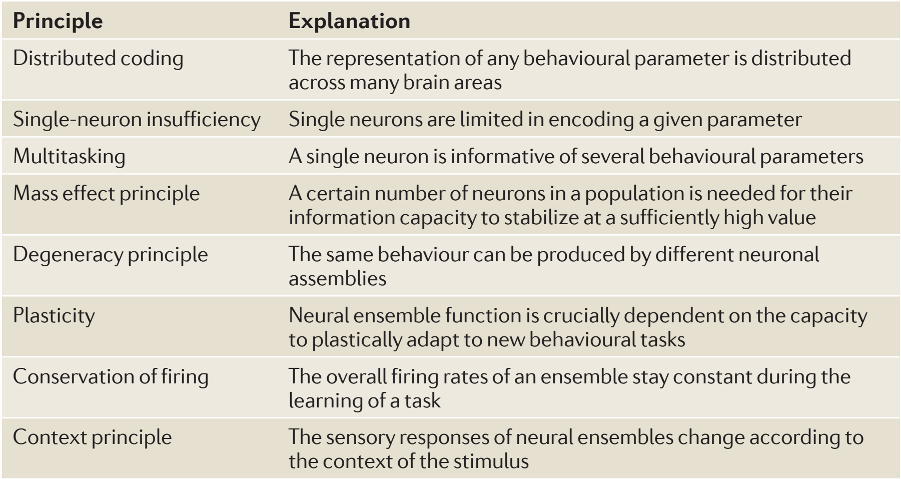
In order to see with clarity how BCI research can contribute to basic neuroscience it is necessary to look in detail at some examples which support my claim. In each case the explanatory knowledge that arises from the experiment involving bionics or prosthetics is not answering a question concerning the layout of an actual neural circuit or mechanism. For this reason it does not matter if plasticity in the circuits has occurred during the experiment, and that the function performed by the mechanism is multiply realizable. Instead, BCI research can answer questions about the operational principles that allow a range of neuronal mechanisms to do what they do.

The first example is from a paper by Carmena and colleagues (2003). They discuss the significance of one of their findings with regards to "the ongoing debate of two opposing views of what the motor cortex encodes" (205). Having observed that tuning depth (roughly, the strength of a neuron's selectivity towards its preferred direction of movement) decreases during the operation of the BCI controlling a robotic arm, they note that this could be taken as evidence for the hypothesis that the tuning of motor cortex neurons is governed by proprioceptive feedback and movement dynamics. However, the observation that tuning depth is still significantly reduced when the monkey uses the BCI while still being allowed to move its real arm lead Carmena and colleagues to the conclusion that the alternative hypothesis, that tuning is governed by abstract motor goals, is also partially true. They argue that this conclusion is supported by the finding that improvement in performance using the BCI is correlated with increases in the tuning depth, suggesting that the motor cortex adapts to the BCI in reformulating motor goals and incorporating visual feedback concerning the operation of the robotic limb or computer cursor, as an alternative to proprioceptive feedback.

Now, the correctness of these specific inferences is not relevant here; what is important is the type of question that these researchers seek to answer. This debate about what the motor cortex encodes does not concern any precise specification of a neuronal circuit, their directional preferences, and patterns of connection. Instead, it asks what general explanation accounts for the tuning properties of these neurons, whether it is movement dynamics or abstract motor goals. Crucially, the answer to this question can be the same even though, as is observed, tuning strengths and preferences change due to the insertion of the BCI. In fact, it is the very observation of the extent and direction of those plastic alterations in neuronal preference that is used to formulate the answer. We have an example of an issue in basic neuroscience that can be addressed with BCI research not only in spite of, but *because of* neuroplasticity. Moreover, this supports the claim, rejected by Craver, that BCI's are a privileged method in certain contexts in basic neuroscience. It is their effectiveness in inducing plastic changes which makes them uniquely useful with regards to certain explanatory goals. Crucially, data gathered from a brain-machine hybrid system, which realizes motor control in an appreciably different way from the brain by itself, can still illuminate mechanisms of motor control in the natural system[19].

We see now that changing the brain can be a way of explaining the brain, if the explanation that is sought is of a general feature of neuronal circuits that remains invariant with plastic modifications induced by the BCI. A recent review by Nicolelis and Lebedev (2009) supplies us with many more examples of this kind. Table 1 lists their "principles of neural ensemble physiology". These are operational principles that determine how neurons work together in the cortex to bring about motor control with or without the bionic implant. It is important to note that most of these have been validated by findings from non-bionic research, while others remain controversial. I will discuss a few of these in detail. Note that Nicolelis and Lebedev do not present their principles as mere "maker's knowledge". They write that:

> "BMI's provide new insights into important questions pertaining to the central issue of information processing by the CNS [central nervous system] during the generation of motor behaviours"



Principle	Explanation
Distributed coding	The representation of any behavioural parameter is distributed across many brain areas
Single-neuron insufficiency	Single neurons are limited in encoding a given parameter
Multitasking	A single neuron is informative of several behavioural parameters
Mass effect principle	A certain number of neurons in a population is needed for their information capacity to stabilize at a sufficiently high value
Degeneracy principle	The same behaviour can be produced by different neuronal assemblies
Plasticity	Neural ensemble function is crucially dependent on the capacity to plastically adapt to new behavioural tasks
Conservation of firing	The overall firing rates of an ensemble stay constant during the learning of a task
Context principle	The sensory responses of neural ensembles change according to the context of the stimulus

Table 1 (from Nicolelis and Lebedev 2009, permission needed)

[19] One might object that this experiment works by intervening on a natural mechanism in the brain, not by modelling the hybrid mechanism as a route towards modelling the brain. I would disagree with this interpretation of the experiment. While the BCI is certainly a tool for intervening on the natural system, my central point is that findings from the hybrid system serve rather straightforwardly as the bases for hypotheses about the natural system. Scientists *are* modelling the hybrid system, but it turns out that coding in the hybrid system need not be characterised any differently from the natural one in spite of cortical reorganisation.

For instance, a central question in theoretical neuroscience has been over whether the brain uses a population code ("distributed coding") to represent perceptual features or to control actions, or a single neuron code featuring the infamous "Grandmother cells" (see Barlow 1972, Kenet et al. 2006). Nicolelis and Lebedev weigh in in favour of the population code, arguing that BCI research has found relatively large populations of around 50 neurons are required to accurately drive a robotic limb[20]. The "Single-neuron insufficiency" and "Mass effect" principles make similar claims regarding the importance of neuronal populations. In all cases, these principles are indifferent to the exact arrangement of neurons in the populations, so that it is irrelevant if a population is multiply realized.

In effect, these "principles" are at a more abstract level than the circuit-level descriptions that Datteri and Craver take to be the aim of the research which they discuss. They do not amount to a typical "how actually" model, though the principles would have to apply to any such model if it were to be built. For that reason it does not matter if the neural realizations change – these principles are applicable to neural behaviour in unmodified and modified systems. It is interesting that three of the principles – "Degeneracy", "Plasticity" and "Conservation of firing" – explicitly refer to what happens as changes occur to the neural systems.

However, in order for these principles to apply across experimental preparations, even when plasticity occurs, it has to be assumed that even if neural circuits undergo plastic modification, the brain does not begin to do things in radically different ways just because of the introduction of the BCI. For example, that it takes roughly the same number of neurons to control a robotic arm as it does to control the real arm. This actually amounts to an endorsement of Datteri's other caveats: ArB1, that there are no uncontrolled perturbations arising due to the implant, and ArB3, that the hybrid system is governed by the same regularity that governs the biological system. So in the end it is fitting to give a positive verdict on two out of Datteri's three assumptions[21].

As it happens, Datteri does also talk favourably about a different kind of hybrid experiment (Reger et

[20] This finding is somewhat controversial as other research groups have reported BCI's operating with fewer neurons being recorded (Serruya et al. 2002, Taylor et al. 2002). Still, it seems that a population code of some sort is in play since no groups advocate a single-neuron code for motor control.

[21] I will discuss this result in the next section.

al. 2000 and Karniel et al. 2005) which, as he puts it, sheds light on the "mechanisms of synaptic plasticity" (322). That is, he apparently does share the insight that plasticity itself is an appropriate target of investigation in neuroscience. Unfortunately, this does not lead him to modify his conception of experimental aims to arrive at a more charitable reading of other bionic experiments that induce plasticity. One option would have been to supplement the idea that the goal of research is a "how actually" model of the target mechanism with the addition that "actually" can be in terms of "what neuron goes where", but also "what rules lie behind the organisation of the neurons". Or to turn again to Craver's schema, in addition to mechanistic, phenomenal and affordance validity, we may add *organisational* validity[22]. This is the test of whether the model is organised in the same way that the biological system is, e.g. using roughly the same size population of coding units, and applying the same rules of coding. As we have seen in this section of the paper, bionic models contribute to basic neuroscience to the extent that they have both affordance and organisational validity.

The question now arises as to whether these additions to Datteri's and Craver's schemas can slot comfortably into their mechanistic account. Are "principles of plasticity" and "mechanisms of plasticity" the same thing? On the one hand, the accounts of plasticity invoked here are not tied to a realisation in a particular neuronal circuit and are not, therefore, typical cases of mechanism description. Still, low level molecular mechanisms of synaptic plasticity might actually be conserved across multiple realizations of the circuit. It is also worth noting that neuroscientists describe their findings as revealing mechanisms even when situated at this fairly high level of abstraction. To take an example of BCI research from the Schwartz laboratory, Legenstein and colleagues (2010) set out to uncover the learning rule behind the modification of motor neurons' tuning preferences in the prosthetic reaching task of Jarosiewicz (et al. 2008). The mathematical model that they use to account for the data (a variation on Hebbian learning) is repeatedly referred to as simulating the "learning mechanism". It can only be that this "mechanism" is multiply realised by different groups of neurons with a variety of tuning preferences, as they adapt plastically to the BCI task. So it remains to be seen if Craver's account of mechanistic explanation, which casts multiple realisation as an "epistemic problem", can in the end incorporate such usages.[23]

[22] Note that in Craver's definition of mechanistic validity, the model's representations of parts, activities, *and* organizational features must all be relevantly similar to the actual mechanism's. The crucial point of this section is that validity with respect to organization can come apart from more anatomical accuracy concerning parts (neurons), and so needs to be evaluated separately. See discussion in Section 4 below.

[23] Here is another example from (non-bionic) visual neuroscience: Freeman et al. (2011) present new fMRI data on

4. *Conclusions and Questions*

In this paper I have argued for two mutually supportive claims, one negative, the other positive. The negative claim is that Datteri's no-plasticity constraint is a methodological norm that is inappropriate not only for BCI research but for systems neuroscience in general. The positive claim is that with an enriched account of the aims of brain research, one which includes organisational principles as a primary target, it is easy to show, contra Datteri, that BCI experiments which induce plasticity can in fact contribute to basic neuroscience; and, contra Craver, that BCI's do have advantages over traditional tools when scientists are addressing certain explanatory problems. Along the way, various questions may have arisen over the possible limitations of some mechanistic approaches in philosophy of science, the normative ambitions of philosophy of science, and the plurality of explanatory goals and evaluative criteria that are called for in the philosophy of neuroscience. In this final section I will address these issues left outstanding, though by necessity my responses here are brief and serve largely as pointers towards further research.

The key finding of section 3.2 was that validity of a model with respect to its representation of parts and activities can diverge from validity with respect to representation of organisational principles. Craver (2010) groups these all together as *mechanistic validity,* whereas I treat mechanistic validity (parts and activities) separately from *organisational validity.* Note also that in my examples organisational validity is inversely correlated with *completeness.* That is, the less complete, and the more abstract and idealised a model is, the better it is able to highlight organisational principles that are invariant with anatomical changes. The contribution of BCI research to basic neuroscience largely comes in the form of the organisational validity of its models and explanations, so it is this failure to distinguish and emphasise this evaluative dimension that has lead Craver and Datteri to underestimate such techniques. These different evaluative dimensions are accompanied by different explanatory goals. The explanatory target of an organisationally valid model is e.g. a principle of neural coding in M1, whereas the explanatory target of a mechanistically valid model is e.g. a realistic description of the

orientation tuning of neurons in primary visual cortex, which they account for in terms of the retinotopic organisation of V1. They write that, "our results provide a mechanistic explanation" (p.4804) of the pattern of findings. Again, what they describe is an organisational principle, rather than a detailed circuit model.

motor circuit for reaching in monkey M1.

Of course, a mechanistic philosophy of neuroscience could be expanded to include the explanatory and evaluative dimensions that BCI research requires. The kind of pluralism I have in mind is compatible, not competitive (Mitchell 2002). For example, it could be said that relative to the circuit description goal, BCI research is not explanatory, but relative to the coding principles goal it is. It should be noted, however, that the "principles" or "learning mechanisms" uncovered by BCI research are not "mechanism sketches" or "how possibly" models, i.e. models whose details are left incomplete so they may be filled in with later research. The finding of a coding rule for motor cortex neurons is a viable endpoint of research in itself. Yet mechanist accounts have tended to treat models lacking in completeness as mere way-stations towards full blown "how actually" models, or else as pragmatically convenient tools that should not be considered as offering satisfactory explanations[24].

Again, the mechanist approach as formulated by Craver and Datteri could be reformulated in order to be more appreciative of models lacking in mechanistic detail, but it ought to be contrasted with some model-based approaches in philosophy of science which emphasise the explanatory virtues of incomplete models (e.g. Cartwright 1983, Batterman 2002), or highlight the instrumental *and* representational value of idealised models developed by scientists for specific tasks (e.g. Wimsatt 1987, Morrison 1998). What is interesting, is that starting out from an approach like this it would be virtually unthinkable to examine BCI research and conclude that the methodological norms employed by the scientists were deficient just because of plasticity and divergence in anatomical detail across different preparations. For on this alternative approach, "a good model is one which doesn't let a lot of these details get in the way" (Batterman 2002: 22)[25].

This brings us to the question of how philosophical accounts of science should be evaluated, and whether divergence from actual scientific practice is sufficient grounds for challenging a normative

[24] See Craver (2010:842) quoted in note 16 above: incomplete models are primarily "useful", and omissions are "sins" rather than explanatory virtues; cf. "How-possibly models are often heuristically useful in constructing and exploring the space of possible mechanisms, but they are not adequate explanations. How-actually models, in contrast, describe real components, activities, and organizational features of the mechanism that in fact produces the phenomenon. They show how a mechanism works, not merely how it might work" (2007:112); and Datteri (2009:308) "Underspecified models and mechanism sketches are progressively refined as model discovery proceeds, until a full-fledged mechanism model is worked out.".

[25] This is obviously a very brief sketch of an alternative approach, which will be presented more fully in a follow up to this paper (Author, in preparation).

claim in the philosophy of science. My case against Datteri in section 2 was essentially making the point that his normative stricture, ArB2, was inconsistent with the practice not only of BCI researchers, but many standard experimental techniques in systems neuroscience. This case does not require any strong assumption that philosophers of science are *never* entitled to make normative claims that are inconsistent with scientific practice. To see how this is so, consider an analogy between philosophical accounts of scientific practice and scientific models representing data sets. Scientific models need not only fit data-sets; they also serve to predict how the data should lie. This predictive function is analogous to a philosopher of science making a normative claim. If data points diverge from the model's prediction, it is not always a good idea to adjust the model to try to fit them. The data can be noisy, and a model which describes all the noise simply makes the mistake of over-fitting. Likewise, if it seems that scientific practice is diverging from ideal practice in random ways, there need be no onus on the philosopher to incorporate such practice in a normative account. However if there is good reason to think that the deviation of data-points is not due to noise, but is a meaningful pattern, it is accepted practice to add a "kludge" – an *ad hoc* modification – to the model to accommodate these anomalies. Furthermore, if another model comes along and can fit the anomalies more or less from first principles (without the need for kludges), then that is reason to take the second model as preferable to the first. Analogously, the addition of organisational validity to Craver and Datteri's mechanistic framework is a kludge, unless it can be shown that the separation between mechanistic and organisational validity can be derived from their "first principles". If not, the alternative model-based approach is to be preferred. Still, the question of whether or not the anomalous data points should be taken as mere noise is ultimately a matter of judgement. In my case against Datteri, I take pains to show that the issue generalises across much of systems neuroscience in order to urge that there is a genuine pattern here, something not to be treated just as noise.

I return, finally, to my tentative endorsement of two out of Datteri's three regulative norms (no uncontrolled perturbations, and sameness of governing regularity), when evaluating organisational validity. One may worry that there will not be a neat separation between cases where plasticity occurs innocuously, and those where gross perturbations occur and governing regularities are broken. If this is so, then it may not be clear in practice if a model of a plastically modified system has achieved organisational validity, or if the other regulative norms have indeed been broken. In response to this issue, I concede that there may in principle be a grey area of cases where the difference between gross perturbation and normal plasticity is not clear, but it is not in this area that BCI experiments are operating. This is because, as mentioned above, the kinds of plasticity occurring in response to the interface are not qualitatively different from those accompanying normal motor skill learning. In other words, all effects lie within the normal operating bounds of motor cortex. It is reasonable, therefore, to assume that the same governing principles apply. Contrast these cases with those of lesion studies in neuropsychology, where the analogous issues arise over the applicability of findings from damaged and reorganised brains to explanation of healthy brains (e.g. Farah 1994, and in the context of developmental disorder see Thomas and Karmiloff-Smith 2002, Machery 2011). In the lesion cases, the kind of plasticity observed involves profound structural reorganisation. It is less likely that the same governing principles are employed in the healthy and lesioned brains, because in the damaged brain different anatomical areas, with different operating norms, may well be employed. Yet even then I would not say that all models of cognitive function derived from lesion studies are problematic just because they cannot meet the regulative criteria for one evaluative dimension (i.e. organisational validity). It may simply be that they require a different evaluative framework.

This paper has asked whether techniques for extending and changing the brain are inimical to the project of explaining the brain, and concluded that they are not. One last point to add is that certain experimental procedures necessarily involve an alteration occurring in the subject matter, yet that does not rule out the validity of the procedure (cf. the measurement problem in physics). It does, however, suggest that there are limits to what can be measured directly, which is a truism, but something often overlooked outside studies in philosophy of science which focus directly on issues of experimental intervention. I have described the complementary nature of neuroscientific methods, given that discovery of one property of a neural systems may come at the cost of knowledge of a related property. A BCI experiment might be ideal for telling you certain things about the motor cortex, e.g. what temporal information in neuronal firing patterns is critical for movement control, yet be ill equipped for resolving a different question, such as the position of motor control maps in natural systems. Fortunately, experimental neuroscience employs an impressive variety of research strategies, each addressing issues that others cannot. It is important that philosophers of neuroscience should recognise the strengths and weaknesses of all of these methods. This is part and parcel of the "mosaic unity" that Craver (2007) aptly describes.

References

Anderson, M. L. (2010). "Neural reuse: A fundamental organizational principle of the brain." Behavioral and Brain Sciences 33: 245–313.

Bach-y-Rita, P. (1972). Brain mechanisms in sensory substitution, Academic Press.

Barlow, H. B. (1972). "Single units and sensation: A neuron doctrine for perceptual psychology?" Perception 1: 371-394.

Batterman, R. W. (2002). " Asymptotics and the Role of Minimal Models." British Journal for the Philosophy of Science 53(1): 21-38.

Bechtel, W. and R. C. Richardson (1993). Discovering Complexity. Princeton, NJ, Princeton University Press.

Carmena, J., M. Lebedev, et al. (2003). "Learningtocontrola brain-machine interface for reaching and grasping by primates." PLoS Biol 1.

Cartwright, N. (1983). How the Laws of Physics Lie. Oxford, Oxford University Press.

Chapin, J., K. Moxon, et al. (1999). "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. ." Nat Neurosci 2(7): 664–670.

Chirimuuta, M. and I. J. Gold (2009). The embedded neuron, the enactive field?. Handbook of Philosophy and Neuroscience. J. Bickle. Oxford, Oxford University Press.

Clark, A. (2004). Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence, Oxford University Press.

Clark, A. (2008). Supersizing the mind: embodiment, action, and cognitive extension, Oxford University Press.

Clark, A. and D. J. Chalmers (1998). "The Extended Mind." Analysis 58(1): 7-19.

Craver, C. F. (2007). Explaining the Brain. Oxford, Oxford University Press.

Craver, C. F. (in press). "Prosthetic Models." Philosophy of Science.

Datteri, E. (2009). "Simulation experiments in bionics: A regulative methodological perspective." Biology and Philosophy 24: 301–324.

David, S. V., W. E. Vinje, et al. (2004). "Natural stimulus statistics alter the receptive field structure of V1 neurons." J. Neuroscience 24: 6991-7006.

de Weerd, P., R. Pinaud, et al. (2006). Plasticity in V1 induced by percpetual learning. Plasticity in the visual system: From genes to circuits. R. Pinaud, L. A. Tremere and P. de Weerd. Berlin, Springer

Dretske, F. (1994). "If You Can't Make One, You Don't Know How It Works. ." Midwest Studies in Philosophy 19(1): 468-482.

Farah, M. J. (1994). "Behavioral and Brain Sciences." Neuropsychological inference with an interactive brain: A critique of the "locality" assumption 17(1): 43-61.

Freeman, J., G. Brouwer, et al. (2011). "Orientation Decoding Depends on Maps, Not Columns." J. Neuroscience 31(13): 4792– 4804.

Ganguly, K. and J. Carmena (2009). "Emergence of a stable cortical map for neu- roprosthetic control." PLoS Biol 7.

Harrison, R. V., K. A. Gordon, et al. (2005). "Is there a critical period for cochlear implantation in congenitally deaf children? Analyses of hearing and speech perception performance after implantation." Developmental Psychobiology 46(3): 252–261.

Hochberg, L., M. Serruya, et al. (2006). "Neuronal ensemble control of prosthetic devices by a human with tetraplegia." Nature 442: 164–171.

Hochberg, L. R., D. Bacher, et al. (2012). "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm." Nature 485.

Hurley, S. (1998). Consciousness in Action. Cambridge, MA, Harvard University Press.

Jarosiewicz, B., S. Chase, et al. (2008). "Functional network reorganization during learning in a brain-computer interface paradigm." Proc Natl Acad Sci U S A 105: 19486 –19491.

Karniel, A., M. Kositsky, et al. (2005). "Computational analysis in vitro: dynamics and plasticity of a neuro-robotic system." J Neural Eng 2: S250–S265.

Kenet, T., A. Arleli, et al. (2006). Are single neurons soloists or are they obedient members of a huge orchestra? 23 Problems in Systems Neuroscience. J. L. v. Hemmen and T. J. Sejnowski. Oxford, Oxford University Press.

Kourtzi Z (2010). "Visual learning for perceptual and categorical decisions in the human brain. ." Vision Res. 50(4): 433-440.

Koyama, S., S. M. Chase, et al. (2009). "Comparison of brain–computer interface decoding algorithms in open-loop and closed-loop control." J Comput Neurosci 29: 73-87.

Legenstein, R., S. M. Chase, et al. (2010). "A Reward-Modulated Hebbian Learning Rule Can Explain Experimentally Observed Network Reorganization in a Brain Control Task." Journal of Neuroscience 30(25): 8400 – 8410.

Legge, G. E. F. and J. M. Foley (1980). "Contrast masking in human vision." Journal of the Optical Society of America 70: 1458-1470.

Lenay, C., O. Gapenne, et al. (2003). Sensory substitution: Limits and perspectives. Touching for Knowing. Y. Hatwell, A. Streri and E. Gentaz. Amsterdam/Philadelphia, John Benjamins Publishing Group.

Leuthardt, E. C., C. Gaona, et al. (2011). "Using the electrocorticographic speech network to control a brain–computer interface in humans." J. Neural Eng 8: 1-11.

Lutz, D. (2011) "Epidural electrocorticography may finally allow enduring control of a prosthetic or paralyzed arm by thought alone." Retrieved 14/4/2012 from: http://news.wustl.edu/news/Pages/21876.aspx.

Machery, E. (2011). Developmental disorders and cognitive architecture. Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory. P. R. Adriaens and A. D. Block. Oxford, Oxford University Press.

Mitchell, S. D. (2002). "Integrative Pluralism." Biology and Philosophy 17(1): 55-70.

Morrison, M. C. (1998). "Modelling Nature: Between Physics and the Physical World." Philosophia Naturalis 35: 65-85.

Musallam, S., B. Corneil, et al. (2004). "Cognitive control signals for neural prosthetics." Science 305: 258 –262.

Nicolelis, M. (2003). "Brain-machine interfaces to restore motor function and probe neural circuits." Nat Rev Neurosci 4: 417–422.

Nicolelis, M. and M. Lebedev (2009). "Principles of neural ensemble physiology underlying the operation of brain–machine interfaces." Nature reviews neuroscience 10: 530-540.

Pinaud, R., L. A. Tremere, et al., Eds. (2006). Plasticity in the Visual System from Genes to Circuits. Berlin, Springer.

Ptito, M., S. M. Moesgaard, et al. (2005). "Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind." Brain 128: 606–614.

Reger, B., K. Fleming, et al. (2000). "Connecting brains to robots: an artificial body for studying the computational properties of neural tissues." Artif Life 6(4): 307–324.

Sagi, D. (2011). "Perceptual learning in Vision Research." Vision Research 51: 1552–1566.

Sanes, J. N. and J. P. Donoghue (2000). "Plasticity and primary motor cortex." Annual review of neuroscience 23: 393–415.

Schirber, M. (2005). "Monkey's Brain Runs Robotic Arm." Retrieved 25/6/2012, from http://www.biotele.com/Monkey.htm.

Schwartz, A. (2007). "Useful signals from motor cortex." J Physiology 579: 581– 601.

Serruya, M., N. Hatsopoulos, et al. (2002). " Instant neural control of a movement signal." Nature 416: 141–142.

Shaw, C. A. and J. McEachern, Eds. (2001). Toward a Theory of Neuroplasticity. Philadelphia, Psychology Press.

Taylor, D., S. Tillery, et al. (2002). "Direct cortical control of 3D neuroprosthetic devices." Science 296: 1829 –1832.

Thomas, M. and A. Karmiloff-Smith (2002). "Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling." Behavioral and Brain Sciences 25(6): 727-750.

Velliste, M., S. Perel, et al. (2008). "Cortical control of a prosthetic arm for self-feeding." Nature 453: 1098-1101.

Wimsatt, W. C. (1987). False Models as means to Truer Theories Neutral Models in Biology. M. Nitecki and A. Hoffman. Oxford, Oxford University Press: 23-55.

Zelenin, P., T. Deliagina, et al. (2000). "Postural control in the lamprey: a study with a neuromechanical model." J Neurophysiol 84: 2880–2887