Minds & Machines (2017) 27:287–313
DOI 10.1007/s11023-016-9406-1

The Epistemic Value of Brain–Machine Systems for the Study of the Brain

Edoardo Datteri[1]

Received: 2 October 2015 / Accepted: 31 October 2016 / Published online: 9 November 2016

Abstract Bionic systems, connecting biological tissues with computer or robotic devices through brain–machine interfaces, can be used in various ways to discover biological mechanisms. In this article I outline and discuss a "stimulation-connection" bionics-supported methodology for the study of the brain, and compare it with other epistemic uses of bionic systems described in the literature. This methododology differs from the "synthetic", simulative method often followed in theoretically driven Artificial Intelligence and cognitive (neuro) science, even though it involves machine models of biological systems. I also bring the previous analysis to bear on some claims on the epistemic value of bionic technologies made in the recent philosophical literature. I believe that the methodological reflections proposed here may contribute to the piecewise understanding of the many ways bionic technologies can be deployed not only to restore lost sensory-motor functions, but also to discover brain mechanisms.

Keywords Brain–machine interfaces · Prosthetic models · Bionic experiments in neuroscience · Robot-based simulation methodologies · Simulations · Discovering mechanisms in neuroscience · Biorobotics · Artificial Intelligence

1 Introduction

Research on brain–computer interfaces (BCIs) is rapidly advancing towards the construction of electronic and robotic systems—sometimes called *hybrid bionic systems*—that may be reliably controlled by the neural activity of living tissues. These technologies may enable restoration of communication, sensory and motor

✉ Edoardo Datteri
edoardo.datteri@unimib.it

[1] "R. Massa" Department of Educational Human Sciences, University of Milano-Bicocca, Building U6, Room 4145, Piazza dell'Ateneo Nuovo 1, 20126 Milan, MI, Italy

abilities lost due to accidents, stroke, or other causes of severe injury (see for example the case of the locked-in patient described in Hochberg et al. 2006). In addition, leading researchers have claimed that bionics technologies can provide unique and new experimental tools to discover brain mechanisms. For example, Wander and Rao (2014) claim that brain–machine interfaces "can … be tremendously powerful tools for scientific inquiry into the workings of the nervous system. They allow researchers to inject and record information at various stages of the system, permitting investigation of the brain in vivo and facilitating the reverse engineering of brain function. Most notably, BCIs are emerging as a novel experimental tool for investigating the tremendous adaptive capacity of the nervous system" (p. 70). Golub et al. (2016) "view BCI as a stepping stone toward understanding the full native sensorimotor control system" (p. 56) and, according to Nicolelis (2003), brain–machine interfaces "can become the core of a new experimental approach with which to investigate the operation of neural systems in behaving animals" (p. 417).

To evaluate whether BCI technologies can live up to these expectations, it is essential to understand how they can be used in neuroscientific research and under what methodological and epistemological assumptions empirical data flowing from bionics-supported experiments can be brought to bear on neuroscientific hypotheses. First steps towards this goal have been taken in (Datteri 2009), in which two bionics-supported methodologies for the discovery of brain mechanisms have been outlined and discussed. Here I will argue that the vast majority of studies reported in the recent scientific literature follow a methodology, called here the *stimulation-connection* methodology, which has not been discussed there. The primary aim of this article is to exemplify (Sect. 2) and analyse some key features (Sect. 3) of this methodology in a contrastive way, that is to say, by comparing it with the *simulation-replacement* methodology discussed in (Datteri 2009).[1]

In Sect. 3.1 I will argue that stimulation-connection and simulation-replacement studies involve structurally similar systems, all being obtained by functionally replacing biological components with artificial devices. In addition, they both use prostheses *qua* functional replacers of biological components in order to test particular neuroscientific hypotheses. The "*qua* functional replacers" clause is not redundant. I will argue that, in some BCI-supported theoretically driven experiments, the artificial part of the hybrid system is not used to replace any biological component. In other cases, the prosthesis actually replaces a biological component, but this fact does not play a crucial epistemic role—in a sense to be discussed—in the neuroscientific discovery process. I will focus on BCI-supported experiments in which brain mechanisms are discovered by functionally replacing biological components with artificial devices—that is to say, in which the artificial device is used *qua* functional replacer. Stimulation-connection and simulation-replacement studies fall in this category.

I will also point out that these two classes of studies differ from one another in a number of aspects. First (Sect. 3.2), they differ in the nature of the scientific

[1] The simulation-replacement methodology is called ArB, from "Artificial replacement of Biological components", in (Datteri 2009). The "simulation-replacement" label is used here to emphasize some of the main differences between it and the "stimulation-connection" method, as discussed later on.

question addressed: the stimulation-connection methodology may assist in the theorization over the biological components *connected* to the prosthesis (hence the "connection" label), while the simulation-replacement methodology may enable one to model the behaviour of the biological component *replaced* by the prosthesis (hence the "replacement" label). Second (Sect. 3.3), they differ from one another in the experimental procedure. The simulation-replacement methodology is akin to the "synthetic method" widely adopted in Cybernetics, Artificial Intelligence, and contemporary biorobotics: theoretical results flow from comparisons between the behaviour of the target biological system and the behaviour of the hybrid system, which can be regarded as a hybrid simulation of the target hypothesis. Stimulation-connection studies make a different, non-simulative use of machine models of biological systems: they apply relatively traditional electrophysiological analysis techniques to neural tissues which are peculiarly stimulated by connection with an artificial device. These distinctions, which will be supported by an analysis of some case-studies, are summarized in Table 1.

In Sect. 4 I will bring the distinction between stimulation-connection and simulation-replacement methodology to bear on some claims recently made by Craver (2010) and Chirimuuta (2013) on the epistemic value of bionic systems. In particular, based on that distinction, I will show that Craver's (2010) arguments, though logically sound, do not support a sceptical view on the role of bionics in neuroscientific research (Sect. 4.1). And in Sect. 4.2 I will argue that some of Chirimuuta's (2013) criticisms to Datteri's (2009) methodological analysis rely on her overlooking the distinction between the two strategies discussed here. Overall, I believe that the piecewise formulation of a taxonomy of bionics-supported experimental methodologies, and the critical analysis of claims made in the philosophical literature on the epistemic value of bionics, may contribute to advancing our understanding of the role of BCI technologies in neuroscientific research.

2 Two Bionics-Supported Studies for the Discovery of Brain Mechanisms

2.1 The Lamprey Reticulo-Spinal Pathway

One of the goals of this article is to outline and discuss the structure of the stimulation-connection bionics-supported methodology for the study of the brain.

Table 1 Summary of the main differences between simulation-replacement and stimulation-connection studies

Method	Focus of inquiry	Experimental procedure
Simulation-replacement	Biological component replaced by the artificial device	"Synthetic method": comparison between target and hybrid system behaviour
Stimulation-connection	Biological component connected to the artificial device	Non-simulative electrophysiological analysis of biological tissues stimulated by connection with artificial devices

The key features of this methodology may be easily identified by comparison with the simulation-replacement experimental strategy discussed in (Datteri 2009), which has been followed by Zelenin et al. (2000) to test a mechanistic model[2] of the lamprey sensory-motor system.

Lampreys are able to maintain a stable roll position by moving tail, dorsal fin, and other body parts in response to external disturbances caused by water turbulence or other factors. A particular portion of the lamprey nervous system—called the reticulo-spinal pathway, *rs* from now on—is thought to play a crucial role in this behaviour. The goal of Zelenin and co-authors is to discover the behaviour of *rs*—more precisely, to discover the relationship between the "input" neurons of *rs* (the reticular neurons) and the roll angles of the animal, which vary as a function of the activity of the "output" spinal neurons. The authors have initially formulated a relatively simple hypothesis r(*rs*) about this relationship. To test it, they have built an electro-mechanical device whose input–output behaviour is r(*rs*). Then they have removed the reticulo-spinal component[3] and replaced it with the electro-mechanical device: the artificial component picked up the activity of the reticular neurons and produced stabilization movements in line with the hypothesized regularity. Finally, Zelenin and colleagues have experimentally tested whether the hybrid system exhibited stabilization abilities comparable to those of the intact system. This has happened to be the case: the authors have therefore concluded that the electro-mechanical device was a good substitute for the *rs* component—and, as a consequence, that the *rs* component actually exhibited the hypothesized input–output regularity r(*rs*).

2.2 Brain Control of Robotic Prostheses in Monkeys

In the study described in (Carmena et al. 2003), two monkeys chronically implanted with micro-electrode arrays in various frontal and parietal brain areas have been trained to perform three kinds of task. In the first one, they had to move a cursor displayed on a screen and reach a target by using a hand-held pole. In the second one, they had to change the size of the cursor by applying a gripping force to the pole. The third task was a combination of the first two. Neural activity was acquired, filtered and recorded during execution of these tasks.

[2] The expression "mechanistic model" is used here to refer to the description of a mechanism (Craver 2007). In what follows I assume that mechanistic models describe the regular behaviour of system components by means of generalizations (Glennan 2005; Woodward 2002). The term "model" is used to emphasize the fact that mechanism descriptions may be more or less *abstract* in the sense clarified by Suppe (1989): they characterize the behaviour of each component as depending on a restricted (though not necessarily narrow) set of factors. For example, a model might characterize the activity of the neurons in a particular brain area as depending only on the firing rate of neurons in another area; a less abstract model would take into account more input or boundary factors. Both models express *counterfactual* generalizations stating that the behaviour of reticular neurons would be such and such, if it depended *only* on that restricted set of factors (Suppe 1989; Woodward 2002). By making these epistemological assumptions I am not claiming that the ensuing methodological discussion of the stimulation-connection methodology is consistent only with this interpretation of the notion of a "mechanistic model".

[3] To be more precise, they have inhibited the activity of this component by using a particular experimental apparatus. See the cited article itself for further detail.

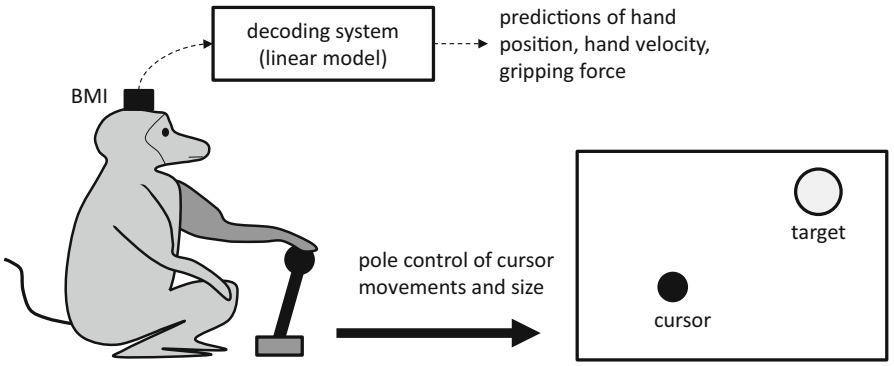

Fig. 1 The experimental set-up in the "pole control" phase

Two different uses have been made of these neural recordings in two distinct phases of the study. During the first "pole control" phase, a reliable correlation has been identified between neural activity and motor behaviour of the monkeys. More precisely, a linear model has been trained to predict various motor parameters—hand position, hand velocity, and gripping force—from brain activity (see Fig. 1).

After obtaining a predictively adequate model, the authors have proceeded with a so-called "brain control" phase. During this phase, cursor position and size were totally disconnected from pole movements: they were instead controlled by the output of the linear model receiving brain activity as input (see Fig. 2). The monkeys had to carry out the same three tasks, obtaining rewards on successful trials.

Notably, in part of the "brain control" trials, neural activity was used to control the movements of a robotic arm and of a gripper located at its tip. Cursor movements and size reflected the movements of the robotic end-effector in space (see Fig. 3), thus providing monkeys with visual feedback on robot movements.

Interesting results with important engineering, therapeutic, and neuroscientific implications have been obtained in these three experimental conditions. A first, basic result, in line with previous studies (see for example Chapin et al. 1999), is that brain control of robotic prostheses is possible. Indeed, after a short learning period, the monkeys became relatively proficient in brain-controlling the cursor, both directly (Fig. 2) and indirectly (Fig. 3). The authors note that, at the very beginning of the "brain control" phase, arm movements were still produced even though they were no more needed to control the cursor. Interestingly, however, after a short period of time, the monkeys ceased to move their limbs while continuing to brain-control the cursor.

Over and above this basic result, which paves the way to important therapeutic applications, the authors have drawn interesting insights on the functioning of the (monkey) nervous system from data obtained during the "pole control" (Fig. 1) and "brain control" (Figs. 2, 3) stages.

Let me start from the "pole control" phase. As pointed out before, at the end of this phase a fairly good model has been obtained, demonstrating that it is possible to

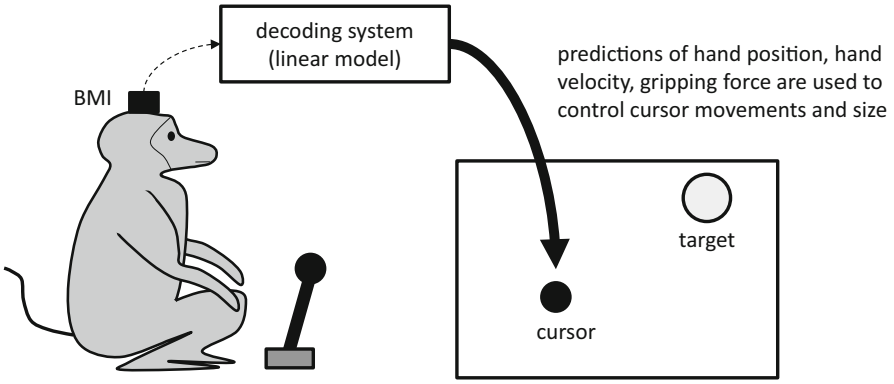

Fig. 2 The experimental set-up in the brain control phase, with the decoder directly controlling the cursor

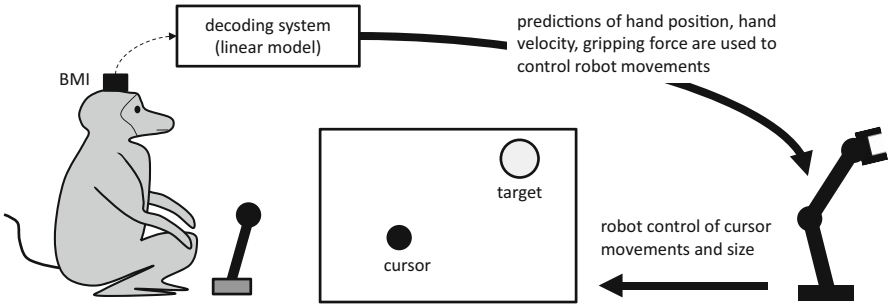

Fig. 3 The experimental set-up in the brain control phase, with the decoder controlling the robot and the robot controlling the cursor

predict various motor parameters from neural activity acquired in frontal and parietal areas with reasonable accuracy. Different brain areas have been found to contribute differently to various aspects of motor behaviour. Moreover, by a so-called "neuron-dropping" methodology (Wessberg et al. 2000), it has been found that the number of neurons required to make good motor predictions based on the linear model changes from area to area (e.g., 33–56 cells in the primary motor area guaranteed accurate predictions of all motor parameters, while 16–19 cells in the supplementary motor area were sufficient to accurately predict hand position and velocity but not gripping force).

The achievement of good performances in the successive "brain control" phase may be taken as indicative of decoder accuracy (even though, as often pointed out in the literature, high brain-control proficiency can be due to the brain's ability to compensate for decoder errors). However, it is worth noting that the *primary* evidential basis on which the claims summarized in the previous paragraph are based—stating that several motor parameters can be predicted from neural activity,

that different areas contribute differently to predicting motor behaviour, and that the quality of prediction depends on population size—flows from data obtained during the "pole control" phase. All these claims concern the predictive value of the model—that is to say, the relationship between model outputs and actual pole-control movements made by the monkeys (see Fig. 1). Information on pole-control movements was clearly available during the "pole control" phase only. Only in this phase it was therefore possible to compare model predictions with pole-control movements.[4] As pointed out before, monkeys rapidly ceased to produce limb movements in the successive "brain control" phase—and, more generally, brain activity in the selected areas ceased to reflect movements of the monkeys' own limbs during the second phase of the study. For this reason, model prediction accuracy could not be sensibly evaluated using data obtained in that phase.

According to the authors, the possibility to simultaneously extract several motor parameters from neural ensembles in various parts of the brain suggest that "motor programming and execution are represented in a highly distributed fashion across frontal and parietal areas and … each of these areas contains neurons that represent multiple motor parameters" (p. 204–205).

Other insightful results obtained in the framework of this study are based on "brain control" data. Some of them flow from control performance analysis. Performances suddenly declined after switching from the pole control (Fig. 1) to the brain control (Figs. 2, 3) mode; however, they progressively improved in successive brain control trials. According to the authors, this result could be explained by assuming that efficient motor control requires a neural representation of the dynamics of the controlled object. At the beginning of the "brain control" phase, the monkeys had to control a totally novel object (a cursor on the screen and a robotic arm): no representation of it could be available in their brains, leading to inefficient motor control. The successive improvements in performance could be explained by hypothesizing that some adaptation process was taking place in the brain, producing a neural representation of the new actuator.

This conjecture is further supported by other results flowing from the analysis of *directional tuning* (DT) *profiles* of individual neurons and ensembles in the "brain control" phase. A DT profile models the relationship between neural activity and direction of movement—for example, by stating that a particular neuron fires maximally whenever the monkey moves her arm leftward. DT profiles have been calculated during the "pole control" and the "brain control" phases, by modelling the relationship between neural firing and cursor movements. Gradual changes in DT profiles have been found during the "pole control" phase. Immediately after switching from pole to brain control, that is to say, at the very beginning of the "brain control" phase, a general decline of DT strength (i.e., of the strength of the correlation between firing activity and movement direction) has been detected. A further decline has been observed when the monkeys ceased to move their limbs. Later on, gradual increases in DT strength have been detected while the monkeys progressively improved their brain-control proficiency, but the levels measured during pole control have been never reached again. The emergence of clusters of

[4] For an example of such a comparison, see the graphs showed in Fig. 2 of the cited article.

neurons with similar DT profiles—that is to say, firing in synchrony with the same movement direction—has been also observed.

According to the authors, these results shed some light on the mechanisms of sensory-motor control in the intact system. The sudden decrease in DT strength after switching from pole to brain control, and especially the fact that DT strength was low even at the very beginning of the second phase when the monkeys were still moving the pole, suggests that DT profiles do not reflect only movement direction *as signalled by proprioception* (this kind of feedback was available at the beginning of the "brain control" phase). The successive increases in DT strength, when proprioceptive feedback was totally uninformative of cursor direction, further support the thesis that monkeys' brains can progressively acquire a neural representation of the movements of the new actuator based on visual feedback only.

> Thus, we hypothesize that, as monkeys learn to formulate a much more abstract strategy to achieve the goal of moving the cursor to a target, without moving their own arms, the dynamics of the robot arm (reflected by the cursor movements) become incorporated into multiple cortical representations. In other words, we propose that the gradual increase in behavioral performance during brain control of the BMI emerged as a consequence of a plastic reorganization whose main outcome was the assimilation of the dynamics of an artificial actuator into the physiological properties of frontoparietal neurons. (p. 205).

Note that this conjecture, supported by data obtained during the brain control phase, consists in a very large-grained, tentative, and incomplete sketch of the mechanism connecting visual feedback to changes in the behaviour of the neurons reached by the interface, that is to say, of the neural mechanism implemented in the

Table 2 Summary of the main results obtained in (Carmena et al. 2003) and of their theoretical implications

Phase	Results and theoretical implications
"Pole control" phase	Model of the relationship between firing activity of cortical neurons and various hand motor parameters
	a) Several motor parameters can be predicted from neural activity
	b) Different areas contribute differently to predicting motor behaviour
	c) The quality of prediction depends on population size
	Motor programming and execution are represented in a highly distributed fashion across frontal and parietal areas, each of which contains neurons that represent multiple motor parameters
"Brain control" phase	Proficiency in controlling cursor movements decreases immediately, and then raises gradually, after switching from pole to brain control
	DT strength immediately decreases after switching from pole to brain control, with a further decrease when the monkeys cease to produce arm movements. Then it starts raising again, without reaching the "pole control" levels
	During brain control of the prosthesis, based on visual feedback, the dynamics of the new actuator become incorporated into multiple cortical representations

biological system to which the prosthesis is connected. This observation will be more extensively justified in the ensuing contrastive methodological analysis of the case-studies reviewed so far. Table 2 summarizes the main results discussed in this section and their theoretical implications.

3 The Stimulation-Connection Methodology: A Comparative Analysis

3.1 Methodological and Structural Similarities

The two studies discussed in the previous section, to which I will refer as "lamprey study" and "monkey study" respectively from now on, share some common methodological features. They both involve *structurally similar* systems, that is to say, hybrid systems obtained by connecting living systems with artificial (computer and robotic) devices. In some parts of both studies, the artificial component of the hybrid system may functionally replace a biological component in performing particular tasks. The robotic device used in the lamprey study functionally replaces the reticulo-spinal pathway plus the motor organs of the animal in the posture control task. And the device used in the "brain control" part of the monkey study replaces the animals' arms and the neural circuitries converting brain activity into efferent commands in the three reaching and grasping tasks. To be sure, what has been called "robotic device" in the last two statements is in fact composed of a brain–machine interface plus a computer-based decoding system generating motor predictions based on neural signals (that is to say, outputting roll angles and hand motor parameters in the lamprey and in the monkey study respectively), and of an electro-mechanical actuator. Let us call "decoder" and "robotic prosthesis" these two parts of the artificial functional replacer (see Fig. 4).

It follows from this description that in the bionic preparations used in these studies one can identify a biological component *replaced* by the artificial functional replacer (the reticulo-spinal pathway plus the motor organs in the lamprey study; the arm plus a cortical decoding circuitry in the monkey study) and a biological

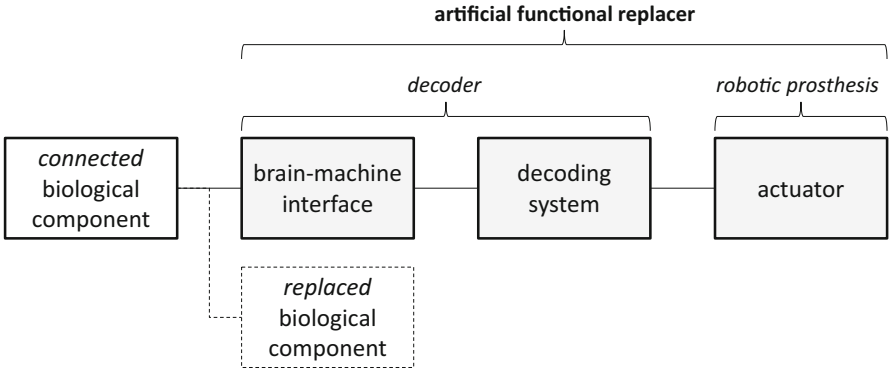

Fig. 4 The structure of a hybrid bionic system

component *connected* to it (that is to say, the "rest" of the system). Note that the replaced component is not physically removed from the intact system in neither of the two studies: it is only made ineffective in performing the task.

The two studies are similar to each other in another methodological aspect. In both studies the artificial device, *qua* functional replacer, plays a crucial experimental role in addressing particular neuroscientific questions. Note the "*qua* functional replacer" remark. In principle, the artificial device (decoder plus robotic prosthesis) or a part of it could be experimentally used in a way that does not rest on its being a functional replacer of a biological component. For example, one may use the brain–machine interface to pick-up neural activity while the monkey is performing a particular task with the control system and the robotic prosthesis turned off. In this way the artificial device would be simply used *qua* monitor of neural activity. Or, one may turn on both devices without searching for correlations between (changes in) brain activity and (changes in) robot control performances.

As a third case, exemplified by the "pole control" part of the monkey study, one may use the brain–machine interface and the decoding system only (with the robotic prosthesis unused or turned off) in order to train the decoding system itself, that is to say, in order to obtain a model able to decode particular motor parameters from brain activity. To this end, one acquires brain activity, generates motor predictions based on the decoding system, compares them with actual movements, and corrects the model implemented in the decoding system so as to obtain better predictions at the next step. The formulation of a model able to predict motor behaviour from brain activity constitutes a striking neuroscientific result, which may have interesting implications for understanding the way movements are represented in, and decoded by, brain circuits. Moreover, such a result would be simply unattainable without automatizing the generation of predictions and the model correction process. It therefore flows from the deployment of innovative machine technologies for the modelling of brain mechanisms. These technological novelties are not devalued by noting that the machine is not used *qua* artificial replacer of a biological component in this case. The fact that the machine used in the "pole control" phase of the study—namely, the interface and the decoding system—was in principle able to drive a functional replacer of the animals' arms has not been crucial to obtain these results. Indeed, there was no artificial replacement at all in this phase of the study: the model was trained with the monkeys using their own limbs to perform the tasks.

The point is not whether the robotic part of the artificial device has been used or not. There is a striking *methodological* difference between the "pole control" phase of the monkey study, on one hand, and the lamprey study and the "brain control" part of the monkey study, on the other hand. Functional replacement has played a crucial epistemic role in the two latter cases. It is by measuring whether the artificial component is an efficient replacer that Zelenin and colleagues have tested their hypothesis on the input–output behaviour of the reticulo-spinal component. And it is by analysing neural activity while the artificial device was functionally replacing the biological arm in the three reaching and grasping tasks that Carmena and colleagues have discovered something important on the mechanisms of plastic change in brain

activity. The artificial devices have been used *qua* functional replacers to obtain these results.

To sum up, the two studies are similar to each other not only in the fact that they both involve a hybrid system, but also in the fact that they have made a theoretically fruitful use of the artificial part of the hybrid system *qua* functional replacer of a biological component. This makes the two studies (among other examples), and in particular the "brain control" phase of the monkey study, not only technologically, but also methodologically novel. It is on the epistemic value of this methodological novelty—that is to say, on the epistemic value of bionic systems used as functional replacers of biological organs—that the ensuing analysis will be selectively focused.[5]

Besides these fundamental analogies, the two studies differ from one another in a number of methodological aspects. The analysis of these differences may enable one to appreciate the novelty of the experimental approach adopted in (Carmena et al. 2003) and in similar studies with respect to the hybrid simulative methodology discussed in (Datteri 2009), and to critically evaluate some arguments recently proposed by Craver (2010) and Chirimuuta (2013).

3.2 Connection Versus Replacement Methodology

As pointed out before, in the two studies reviewed here a biological component is *connected to* an artificial device which may *replace* another biological component. A first methodological difference between the two studies concerns whether the focus of inquiry is on the replaced or the connected biological component. The lamprey study aims at modelling the behaviour of the replaced component. On the contrary, the "brain control" part of the monkey study offers insights to theorize on the neural mechanisms implemented in the component to which the artificial device is connected. Let me elaborate on this point.

The question addressed in the lamprey study concerns the input–output behaviour of the reticulo-spinal pathway controlling roll posture, which is the biological

[5] The distinction made here can be brought to bear on the question of what exactly is a *bionics*-supported neuroscientific experiment. It may be defined as a neuroscientific experiment exploiting in vivo connections between living systems and artificial devices. This definition would be too liberal, however, as traditional neurophysiological experiments—e.g., voltage clamp experiments—would fit it. On the contrary, it would be too restrictive to include in the class of bionics-supported experiments only those experiments in which the target living system controls *robotic* devices, as this would exclude experiments in which the subject brain-controls virtual devices (as, for example, in the set-up illustrated in Fig. 2). Then, one may include in that class all and only those neuroscientific experiments which involve hybrid systems whose structure can be described as in Fig. 4. It is worth noting, however, that the label "hybrid bionic system" is typically used in the contemporary scientific literature to refer to systems in which the artificial device functionally replaces a biological component. That is to say, contemporary bionics research is focused on the realization of devices which are *essential* for people suffering from motor or sensory limitations to perform particular tasks, as they can replace sensory or motor organs. For this reason, here I will restrict the label "bionics-supported neuroscientific experiment" to all and only the experiments which make an essential use of artificial devices *qua* replacers of biological components. The label therefore does not apply to the experiments carried out in the "pole control" phase of the monkey study, even though they have involved a system that could be structurally described as in Fig. 4. As suggested by the analysis made in this article, overlooking this distinction may obscure the methodological novelty of the experiments carried out in the "brain control" phase of the monkey study.

component replaced by the artificial device. As discussed above, the authors' strategy consists in replacing the target component with an artificial device whose input–output behaviour is known, and checking whether the hybrid system can produce the same behaviour as before. In that case, one may be induced to conclude that the input–output behaviour of the replaced component matches the behaviour of the replacer—hence, since the behaviour of the replacer is known, one would obtain a description of the former. Bionics-supported experiments and methodologies in which functional substitution with an artificial device enables one to inquire into the behaviour of the replaced biological component will be called *replacement* experiments and methods from now on.

Part of the monkey study—the "pole control" phase—has been devoted to developing a model of the relationship between firing activity of cortical neurons and various hand motor parameters, that is to say, a model of the behaviour of the biological component that in the "brain control" phase has been functionally replaced by the artificial device. In the previous section I have argued that the machine has not been used *qua* functional replacer in the first part of the monkey study. On the contrary, in the "brain control" phase the authors have analysed changes in the directional tuning of cortical neurons while the monkeys were learning to brain-control the prosthesis, and have brought these results to bear on what happens in the brain after connection with an artificial functional replacer. These results can be brought to bear on the mechanisms enabling the brain to adapt to new tools and, more generally, to develop good motor control abilities (recall the theoretical considerations summarized at the end of Sect. 2.2). In this part of the study, the authors have made experimental use of the artificial device *qua* functional replacer to address scientific questions on the biological component *connected* to the artificial device itself. Unlike the lamprey study, this part of the monkey study can be classified as a *connection* bionics-supported study.

This thesis can be further supported by reflecting on the relationship between the input–output behaviour of the artificial device and of the replaced biological component in the lamprey study and in the "brain control" part of the monkey study. At the beginning of the lamprey study, the authors *do not know* whether the two components have the same behaviour or not: this is exactly the question that the authors want to address. At the beginning of the "brain control" part of the monkey study, on the contrary, the authors *already know* that the decoding system's behaviour matches to a reasonable extent the input–output behaviour of the biological component which is now functionally replaced by it: this was exactly the goal of the previous part of the study. One should be careful to note that other connection studies have tested the ability of the brain to adapt to decoders which were known to be *predictively inaccurate*, as they had been obtained by shuffling the weights of a previously trained model (Ganguly and Carmena 2009 is a case in point). The point is not whether the replacing device's behaviour matches the replaced component's behaviour or not. It rather concerns whether, at the beginning of the study, one *has information* on whether the two behaviours match or not. Such information was unavailable at the beginning of the lamprey study, consistently with my claim that the goal of the study was exactly to discover the behaviour of the replaced component. On the contrary, such information was available at the

beginning of the "brain control" monkey study and of (Ganguly and Carmena 2009), consistently with the claim that the goal of the study was *not* to discover the behaviour of the biological component replaced by the decoding system. I also take for granted that the monkey study was not aimed at modeling the behaviour of the biological component replaced by the robotic prosthesis, i.e., the arm. Therefore, since the artificial system used in that study was composed by a decoding system and by a robotic prosthesis (Fig. 4), and that neither system was the focus of inquiry, one may legitimately conclude that the "brain control" monkey study was aimed at modeling the behaviour of the biological component connected to the artificial system.

A wide variety of bionics-supported experiments of great neuroscientific interest have been performed since (Carmena et al. 2003). Indeed, there are good reasons to claim that the number of bionics-supported connection studies that are reported in the literature is far higher than the number of replacement studies. For example, Ganguly et al. (2011) have claimed that, although previous studies have discovered modifications in neural activity correlated with improvements in brain control performance, "it remains difficult to place such modifications in the context of the large cortical network for motor control" (p. 662). To proceed towards this goal, they have studied the relationship between changes in the activity of the neurons more directly involved in prosthetic control and changes in the activity of what they have called "indirect" neurons, that is to say, neurons reached by the interface but less involved in prosthetic control. Based on that, they have speculated on the functional role that these indirect neurons might have in determining the activity of the "direct" neurons, and on the possible layout of the cortical mechanisms involved in forming internal representations of external limbs and tools. The goal of this study was to inquire on what happens upstream of the neurons more directly involved in motor control, and in particular, to develop hypotheses on the brain mechanisms underlying plasticity (see also Koralek et al. 2012).

The claim that bionics can assist in the study of the biological components connected to the machine is often more or less explicitly made in the scientific literature. For example, Golub et al. (2016) point out that bionic systems can help one to "obtain a more complete understanding of the cognitive processes underlying sensorimotor control" (p. 53) and "to understand how different sensory modalities contribute to sensorimotor control" (p. 55): "because of the similarity of the cognitive processes and brain areas involved in native motor and BCI control, we view BCI as a stepping stone toward *understanding the full native sensorimotor control system*. The BCI paradigm, being a reduced system, offers vastly improved accessibility and manipulability, without simplifying away the complexities of *brain processing that we wish to understand*" (p. 37, emphasis added). It follows from this claim that the focus of inquiry in bionics-supported experiments is the non-simplified part of the brain, that is to say, the biological components connected to the BMI. Similarly, Orsborn and Carmena (2013) point out that "BMI also allows observation of brain areas not directly contributing to the task. What occurs in *other parts of the motor cortex* as a subject learns a neuroprosthetic skill?" (p. 4, emphasis added). According to Wander and Rao (2014), various results flowing from motor bionics-supported experiments have suggested that, during adaptation to a novel

tool, "the brain [is] dynamically modifying internal networks to dissociate changes in neural activity from the motor movements with which they were originally correlated" (p. 71) and that, "even when effective control of the BCI only explicitly requires modulation of activity in a small cortical region, frontal and parietal cortical regions are strongly task modulated during initial performance of the task and less so after extensive training" (p. 72). All these claims point to the role of BMIs in modelling the processes occurring in the part of the hybrid systems connected to the machine. Even more explicitly, the same authors state that "BCIs can be a powerful tool *for scientific inquiry into the very system with which they interface*. BCIs afford the experimentalist opportunities not only to *observe sensorimotor transformations as information travels through the brain*, but also to modify the nature of these transformations in real-time. Most importantly, BCI technology provides a scaffold for scientific experimentation that enables investigation of the nervous system doing what it does best: incorporating new information and rapidly adapting to new constraints" (p. 73, emphasis added).

Let me sum up by mathematical analogy. The behaviour of the replaced components is the unknown variable in the equation the authors of the lamprey study have tried to solve. In the "brain control" part of the monkey study, on the contrary, the behaviour of the replaced component has been fixed by a prosthetic implant which imposes a known mapping from brain activity to movements, in order to simplify the identification of the unknown variables representing brain mechanisms.

In the next section I will argue that the lamprey and the "brain control" part of the monkey study differ from each other also in the nature of the empirical basis which is brought to bear on the theoretical hypothesis under scrutiny.

3.3 Stimulation Versus Simulation Methodology

Machines have been often used in cognitive science and neuroscience as simulations of theoretical models of animal behaviour. The simulative methodology, sketched by pioneers of Cybernetics Arturo Rosenblueth, Norbert Wiener and Julian Bigelow in (Rosenblueth and Wiener 1945), proceeds as follows. Let M be a model of the (cognitive or neural) mechanism hypothesized to produce behaviour B by living system S in particular experimental conditions C. To test if this hypothesis is true, one can build an artificial system A which implements M, and compare A's behaviour in conditions C with B. Behavioural similarities between A and S may induce one to conclude that M can produce B in C. Otherwise, one may be induced to reject the conjecture. Several examples of this methodology can be found in pre-cybernetic, mechanistic investigations on animal behaviour and learning (Cordeschi 2002), in Cybernetics, Artificial Intelligence (Newell and Simon 1961; Simon and Newell 1962), and contemporary biorobotics (Floreano et al. 2014; Datteri and Tamburrini 2007; Tamburrini and Datteri 2005; Webb 2001).

A notable feature of the simulative method is that theoretical conclusions on the relationship between theoretical model M and living system S are obtained by

comparing S' behaviour with the behaviour of the simulation. The fact that computers implementing particular models of heuristic means-end analysis proved able to solve problems with performances comparable to those exhibited by human beings engaged in the same tasks was the main evidential basis used by Newell and Simon to support their theories on problem solving. The fact that the robot described in (Grasso et al. 2000) consistently failed to reproduce lobsters' goal-seeking behaviour was the main evidential basis used by the authors of the study to reject their hypothesis on lobster chemiotaxis. Note that the simulative methodology does not require one to focus on the motor or verbal output of the simulation only. Often, internal comparisons between living and simulation systems are carried out. For example, one may check not only if the output of an artificial neural network reproduces the output of a biological network, but also if the patterns of spiking activity of individual artificial neurons match those of their biological counterparts (see Chou and Hannaford 1997, for an example). Newell and Simon famously asked their subjects to produce verbal reports of their reasoning processes during problem-solving tasks in order to compare their cognitive dynamics with the representational transformations made by the simulation system.

As discussed in (Datteri 2009), the lamprey study exemplifies a hybrid variant of the simulative strategy. While classic, non-hybrid simulation studies typically start with a description of M specifying the input–output behaviour of all the components of the hypothesized mechanism and their organization, the authors of the lamprey study start with a hypothesis on the input–output behaviour of just one component—the reticulo-spinal pathway plus the motor organs of the animal. While in non-hybrid simulation studies one builds a fully artificial system implementing M, the authors of the lamprey study have built an artificial component implementing the input–output behaviour of the target biological component only. More crucially, as in classic simulation studies of Artificial Intelligence, Cybernetics, and biorobotics, the primary evidential basis which is brought to bear on the target hypothesis flows from the analysis of whether the hybrid system's posture maintenance behaviour matches the behaviour of the intact lamprey or not. For these reasons, I take the lamprey study as a *simulation* bionics-supported study.

The "pole control" phase of the monkey study conforms to the simulation methodology in some of the aspects discussed here, as the primary evidential basis to conclude that the decoding system predicts hand motor parameters from brain activity comes from the analysis of whether the monkey's behaviour matches decoder outputs or not. Interestingly, the successive "brain control" phase of the monkey study does not conform to the simulative methodology. The experimental results obtained in this part of the study (see Table 2) flow from an analysis of the relationship between neural and motor parameters of the system (for example, between firing rate and movement direction), and of changes in this relationship, while the animal is learning to brain-control the artificial replacer. In other words, even though an artificial model of the replaced biological component is used, these results are obtained by applying relatively traditional electrophysiological analysis techniques in a radically novel experimental setting—that is to

say, while the brain is stimulated by the presence of a new tool.[6] For this reason, I consider the "brain control" part of the monkey brain as an example of a *stimulation*, non-simulative bionics-supported methodology for the discovery of brain mechanisms.

This is not to say that comparisons between motor control proficiency in the hybrid and in the intact monkey play no theoretical role at all. The achievement of high brain control proficiency is crucial to interpret the detected plastic changes as related to the successful "internalization" of the dynamics of the new tool. But the outcome of this comparison is not the main empirical basis on which the theoretical conclusions summarized above are made to rest, as in the (hybrid) simulation methodology: they crucially flow from a neurophysiological analysis not intended to provide data for output or "internal" comparisons. DT analysis and the neuron-dropping procedure adopted in the "brain control" phase are not parts of the "synthetic method".

To sum up, the "brain control" phase of the monkey study exemplifies how machine models of biological components can be used in a way that is interestingly different from the way machine models have been traditionally used in cognitive science and neuroscience. The distinction sketched here between simulation and stimulation bionics-supported experiments, as well as the distinction between connection and replacement experiments, will be used in the next section to evaluate some claims made in Craver (2010)'s and Chirimuuta (2013)'s analyses of the epistemic role of bionics.

4 Underdetermination, Plasticity, and the Methodological Requirements of "Good" Bionic Experiments

4.1 Craver on the Epistemic Role of Bionic Experiments

In his article, Craver (2010) focuses on "what, if anything, building a prosthetic mechanistic model adds to our confidence that we have a valid mechanistic model over and above the degree of confidence provided by models and simulations alone"

[6] To be sure, experiments in electrophysiology often involve artificial stimulations of the target biological tissue. For example, one may intervene on the membrane potential of particular neurons after blocking specific kinds of ion channels in order to find the threshold above which action potentials are generated in those conditions. Note that, in experiments of this sort, the nature and magnitude of the "input" stimulation (e.g., of changes in membrane potential) do not systematically depend on the effects of that stimulation (e.g., on whether action potentials are generated or not). The "input" parameters are independent of the "output" of the interventions: the experimenter explores a relevant portion of the "input" space and measure the effects in order to find a correlation. Quite on the contrary, in non-simulation, connection bionic experiments, the nature and magnitude of the stimulation received by the biological system crucially depends on the "output" of biological activity—that is to say, on the behaviour of the prosthesis as determined by the biological system itself. Plastic changes in the subject's brain depend on the feedback informing the subject about the way her brain is moving the prosthesis. Brain activity determines prosthetic behaviour; information on prosthetic behaviour determines changes in brain activity. Such a "circular" connection between the nature and magnitude of the stimulations applied to brain circuits and the nature and magnitude of the effects of those stimulations is not established in traditional electrophysiological intervention experiments.

(p. 843). His answer is a qualified "nothing". However, at the end of the paper he draws a stronger conclusion, defending a sceptical view as to whether "the ability to build a successful prosthesis counts as evidence that one knows how the system works" (p. 850).

Craver's argument runs as follows. He refers to a prosthesis as *affordance valid* "to the extent that the behaviour of the simulation could replace the target in the context of a higher-level mechanism" (p. 842). A robotic arm enabling one to perform all the movements and actions she could perform with a biological arm—such as grasping, lifting, or pushing objects—is affordance valid. A prosthesis is said to be *phenomenally valid* to the extent that "its input–output relationship is relevantly similar to the target input–output relation" (p. 842) under standard or non-standard conditions. A robotic arm, for example, is phenomenally (or behaviourally) valid if the relationship between its inputs and movements is relevantly similar to the relationship between the inputs and the movements of the biological arm it is replacing. Finally, a prosthesis is said to be *mechanistically valid* to the extent that its parts, activities, and organizational features are relevantly similar to the parts, activities, and organizational features of the target system (p. 842). A robotic arm, for example, is mechanistically valid if its internal mechanism is relevantly similar to the internal mechanism governing the replaced arm (mechanistic validity will be discussed more in detail below).

Craver's first point is that affordance valid prostheses need not be phenomenally valid. He rightly argues that a robotic system can partially replace the functionality of a missing arm or leg even if its input is very different from the input of the replaced biological component. The arm prosthesis described in Carmena et al. (2003), for example, is affordance valid (the monkeys could use a brain-controlled prosthesis to carry out tasks which they had previously carried out by using their own hands). However, the neurons whose activity controlled the prosthesis were not those providing input to the animals' biological arms, as the prosthesis was controlled by the activity of various frontoparietal neural ensembles acquired through a multi-electrode brain–machine interface. Even though these brain areas participate in motor control, monkey arms are not *directly* connected to these areas. The input of the prosthesis was very different from the input of biological monkey arms. For this reason, the input–output behaviour of the prostheses was radically different from the input–output behaviour of monkey arms, and the prosthesis was therefore to be regarded as *phenomenally invalid*.

The fact that affordance validity does not imply phenomenal validity has, according to Craver, an important consequence as to whether bionics research can really contribute to the discovery of brain mechanisms: being able to build an affordance valid prosthesis does not imply having understood the input/output relationship characterizing the replaced biological component.

Another point made by Craver is that affordance and phenomenal validity do not imply mechanistic validity: "a prosthetic model might be affordance valid and phenomenally valid yet mechanistically invalid"; "building a functional prosthesis that simulates a mechanistic model is insufficient to demonstrate that the model is mechanistically valid" (p. 845). Recall that a prosthesis is mechanistically valid to the extent that its parts, activities, and organizational features are relevantly similar

to the parts, activities, and organizational features of the target. Therefore, here Craver is arguing that a prosthesis may be affordance or phenomenally valid even though its internal mechanism is not relevantly similar to the internal mechanism of the target. This is because many different mechanisms, in principle, could produce the same input–output behaviour: in the philosophical jargon, the mechanism is *underdetermined* by the device's input–output behaviour, or, equivalently, the latter is said to be *multiply realizable*.

> Phenomena are *multiply realizable* in lower-level mechanisms. Multiple realizability obstructs the inference from a model's phenomenal validity to its mechanistic validity. The space of phenomenally adequate simulations might well be too large and heterogeneous to provide any assurance that the mechanistic features of a phenomenally adequate simulation are relevantly similar to the mechanistic features of the target (p. 844).

It is important to understand what exactly Craver means with "target" in these claims. The problem of multiple realizability obstructs inferences from the behaviour of a system to the mechanism producing that behaviour. In a simulation study, in particular, one builds an artificial system A simulating a theoretical model of target (living) system S (see Sect. 3.3). In this context, the problem of multiple realizability can be stated as follows: the fact that simulation system A reproduces the input–output behaviour of target system S is insufficient to conclude that A's internal mechanism is relevantly similar to the mechanism governing S. Now, Craver warns us that building a prosthesis that reproduces the input–output behaviour of a biological component (thus being phenomenally valid) is insufficient to conclude that the prosthesis reproduces the internal mechanism *of that biological component*, namely of the component *replaced* by the prosthesis. His point is that a phenomenally valid arm prosthesis, for example, can fail to reproduce the mechanism governing the replaced arm. To sum up, Craver argues that being able to build a phenomenally valid prosthesis is not sufficient to demonstrate that one has understood the mechanism governing the biological component replaced by the prosthesis.

Taken together, the arguments discussed so far allow Craver to conclude 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" (p. 850). Accurate replication of the input–output behaviour and internal mechanism of the replaced component is not needed to build an efficient prosthesis (this is the blessing for engineers); at the same time—Craver argues—one can build an efficient prosthesis without having understood the behaviour of the replaced component and its internal mechanism (this is the curse for basic researchers). The sceptical theses on the epistemic value of bionic systems introduced above flow, in Craver's view, from this conclusion.

Craver's claims that affordance validity does not imply phenomenal validity, and that phenomenal validity does not imply mechanistic validity of a prosthesis, are logically correct. However, based on the distinction between simulation-replacement and stimulation-connection methodologies made above, I will argue that they

do not provide good reasons to believe that bionic systems are of little help in discovering brain mechanisms.

Craver's first point is that affordance valid prostheses need not be phenomenally (behaviourally) valid. This thesis, interpreted as a general rule, is true: one can find many examples of animals learning to control robotic prostheses whose input–output behaviour does not match the behaviour of the replaced components (as Craver rightly points out, efficient control can be achieved even when the input of the prosthesis is very different from the input of the replaced component). It is worth noting, however, that *in the framework of particular studies*, an example being the lamprey study, it may be safe to infer phenomenal validity from affordance validity. Craver claims that bionic prostheses are never connected to the original brain inputs of the replaced biological component ("no currently available [bionic] device, to my knowledge, makes use of just those brain inputs that move limbs in typical animals", p. 845). This is not true: the electro-mechanical device in the lamprey study described above was driven by the same neural input (i.e., by the activity of the reticular neurons) that, in intact lampreys, drive the behaviour of the target reticulo-spinal component. And the fact that the prosthesis proved able to functionally replace the target component (thus being affordance valid) was taken as a basis to conclude that its input–output behaviour matched the target component's input–output behaviour (i.e., that it was phenomenally valid too). Another study in which the prosthesis is driven by the same inputs that drive the behaviour of the replaced component is reported in (Le Masson et al. 2002).

The fact that, in the framework of particular studies, it is safe to infer phenomenal validity from affordance validity does not invalidate Craver's thesis that affordance valid prostheses need not be phenomenally valid, interpreted as a general rule. It only suggests that Craver's thesis should not be improperly taken as supporting a too sceptical thesis on the epistemic value of prosthetic models. In some cases, having built an affordance valid prosthesis implies having understood the input–output behaviour of the replaced component.

Another point made by Craver is that affordance and phenomenal validity do not imply mechanistic validity: being able to build an artificial device which reproduces the input–output behaviour of a biological component does not imply that one has understood the mechanism governing it. It is important to clarify what exactly we can learn from this claim. Evidently, as argued before, Craver is warning us that building a working (affordance or phenomenally valid) prosthesis does not imply that we have understood the mechanism governing the behaviour *of the replaced biological component* (for example, of the replaced arm, in the case of an arm prosthesis). His argument, therefore, specifically concerns the epistemic value of bionic experiments in which the target component (i.e., the component whose behaviour or internal mechanism is to be discovered) is replaced by the prosthesis (i.e., those experiments referred to in this paper as *replacement* experiments). Moreover, it is intended to support a sceptical view as to whether such experiments can assist in discovering *the internal mechanism* of the replaced component.

I will reply to Craver's argument in two steps. First, as borne out by many successful model-oriented simulation studies, the fact that the input–output behaviour underdetermines the internal mechanism does not imply that building a

phenomenally valid prosthesis provides *no evidence at all* regarding the structure of the replaced component's internal mechanism. Second, as suggested earlier, bionic technologies enable experimental strategies that are significantly different from the simulative replacement methodology described in Sect. 2.1, and which contribute to the discovery of brain mechanisms by shedding light on the input–output behaviour of the replaced component rather than on the replaced component's internal mechanism. Craver's multiple realizability argument does not apply to these strategies, and therefore does not support a generalized sceptical view of the epistemic value of bionic experiments.

Let me begin from the first step. Craver is right in pointing out that, strictly speaking, phenomenal validity does not imply mechanistic validity. However, this argument does not exclude that building a phenomenally valid prosthesis can *provide evidence* contributing to the discovery of the internal mechanism of the replaced component. Craver construes evidence as "a finding that shapes (or constrains) the space of possible mechanisms for a given phenomenon" (Craver 2010, p. 843). Now, the fact that the input–output behaviour of the prosthetic component matches the input–output behaviour of the replaced component—thus, that the prosthesis is phenomenally valid—is not a conclusive reason to claim that the prosthesis' internal mechanism is relevantly similar to the replaced component's internal mechanism. Nevertheless, it is a *good* reason to include the prosthesis' internal mechanism within the space of the possible mechanisms governing the replaced component. Similarly, realizing that the prosthesis is not phenomenally valid may be legitimately taken to be a good reason to exclude its internal mechanism from the space of how-possibly models of the replaced component. Purely behavioural tests often lead Artificial Intelligence and biorobotics researchers to expand or restrict the space of the possible models of the target living system. An example is the model-oriented biorobotic study reported in (Grasso et al. 2000), in which mismatches between the overt behaviours of the simulation and of the target system—with no "internal" comparison between the two systems—were taken as reasons to conclude that the mechanism implemented in the robot could not be a good model of the animal. To sum up, Craver is correct in claiming that phenomenal validity does not *imply* mechanistic validity. Nevertheless, his underdetermination argument does not exclude that phenomenal validity can *provide evidence* constraining the space of the possible models of the target system—thus, that it can contribute to model discovery.

The second step is more easily taken. Craver's underdetermination argument focuses on the fact that it is mistaken to infer a prosthesis' mechanistic validity from its phenomenal validity. Recall that a prosthesis is mechanistically valid to the extent that it realizes the mechanism governing the *replaced* component. Therefore, Craver's underdetermination argument supports a sceptical view as to whether building a phenomenally valid prosthesis can contribute to discovery of the mechanism governing the *replaced* limb. However, as discussed before, in bionic *connection* studies (for example, in the "brain control" phase of the monkey study) one theorizes on the behaviour of a biological component by *connecting* it to, rather than by replacing it with, artificial devices. The underdetermination argument has no force against this kind of studies. Consider also that the argument applies only when

the mechanism governing the behaviour of the replaced component is inferred from the behaviour of the prosthesis or the behaviour of the hybrid system only. Inferences of this kind are made in what have been called here *simulation* bionic experiments. As argued in Sect. 3.3, in *stimulation* bionic studies neural mechanisms are inferred not only from the analysis of overt hybrid system behaviours, but also from the analysis of neural activity while the subject is learning to control the prosthesis. The underdetermination argument, as proposed by Craver, has therefore nothing to say on experiments of the latter kind—thus, among other examples, on (Carmena et al. 2003).

One should also be careful to note that Craver's underdetermination argument concerns inferences from the replaced component's behaviour to its internal mechanism. That is to say, his argument is intended to support a sceptical view as to whether bionic technologies can assist in discovering *the internal mechanism of a component* of a sensory-motor mechanism. But it does not rule out the possibility that bionic technologies can assist in discovering the *input–output behaviour of a component* of a sensory-motor mechanism—a discovery that may be of theoretical interest with respect to the broader goal of discovering the structure of the mechanism governing the containing system. For example, the goal of (Zelenin et al. 2000) was to identify the behaviour of the reticulo-spinal component, and not its internal mechanism. Nevertheless, by achieving this result, they have contributed to the modelling of the whole lamprey sensory-motor system. Craver's underdetermination argument does not rule out contributions of this sort.

In sum, Craver (2010) does not offer strong arguments for excluding that bionic technologies can provide evidence for the discovery of brain mechanisms. In particular, his arguments—though logically correct—apply neither to stimulation-connection experiments, nor to simulation-replacement experiments aimed at identifying the behaviour of the target component and not its internal mechanism (as in the lamprey study).

To conclude, another claim made by Craver in his article is that bionic methodologies do not offer significant epistemic advantages over other kinds of methodologies currently adopted in neuroscience: "the effort to build a prosthetic model allows a decisive test of affordance validity but offers no distinct advantages for assessing the model's phenomenal and mechanistic validity" (p. 841). However, as acknowledged by Craver himself, simulations in general can significantly speed up the process of evaluating the behavioural implications of the mechanistic model at issue. In some cases, this is likely to be also true for *bionic* devices. Consider the lamprey study. To assess whether the reticulo-spinal pathway performed the hypothesized regularity, the authors could well have recorded reticular activity and measured roll movements in an intact, swimming lamprey, in search of a correlation between the two. Even though it is difficult to say a priori whether building such a recording apparatus would have been more difficult than setting-up the bionic preparation described in the article, the bionic solution offered a relatively direct means of assessing whether the authors' hypothesis was correct: instead of searching for a correlation, they just checked if the hybrid animal was able to stabilize itself. And consider the stimulation-connection experiments described in Sect. 2.2. It is reasonable to believe that connection with an actual prosthetic device

played a crucial role in the development of the various theoretical hypotheses on the neural assimilation of external tools reported in (Carmena et al. 2003; Lebedev et al. 2005; Nicolelis 2011)—and therefore, that the deployment of bionic technologies has made a decisive contribution, vis-à-vis other experimental methodologies, to achieving these results. There are no sufficiently general criteria for assessing the relative epistemic advantages of vastly different (bionic versus non-bionic) methodologies. However, one may reasonably believe that at least in some cases (as in the examples discussed here) bionics can offer particularly insightful and informative experimental methodologies for the discovery of brain mechanisms.

4.2 Chirimuuta on the Methodology of Bionics-Supported Experiments

Chirimuuta (2013) has argued that bionic technologies may play an important role in neuroscientific research: in her words, "*changing* the brain can help in the project of *explaining* the brain" (p. 614). In particular, she stresses that connection with an artificial replacer can help one theorize over the mechanisms of brain plasticity. The analysis of the lamprey and monkey studies carried out in this article provides many reasons to agree with her that changing the brain by connecting it with an artificial device which replaces functionally a component of the target system can assist in the mechanistic explanation of the target system's behaviour. However, I will argue that some arguments proposed by Chirimuuta on the methodology of bionics-supported experiments (more specifically, some criticisms to particular methodological claims made by Datteri 2009) rest on the assumption that studies like (Carmena et al. 2003) are instances of a simulation-replacement methodology, contrary to what I have argued before. In this section I set out to apply the distinction between simulation-replacement and stimulation-connection studies proposed above to challenge these arguments and to discuss the methodological requirements of "good" bionics-supported experiments.

As earlier recalled, the methodology of simulation-replacement experiments has been extensively discussed in (Datteri 2009). There it is argued that, in order to infer the behaviour of the replaced component from the behaviour of the overall system, no plastic adaptation process must occur in the non-replaced, biological portion of the system. This constraint is labelled by Chirimuuta as the "no-plasticity constraint". It is easy to understand why this constraint is to be placed on simulation-replacement experiments by reasoning on the methodology of the lamprey experiment. As pointed out above, in that study the authors have functionally replaced a biological component with an artificial device whose behaviour is known. The fact that the hybrid system has proven able to replicate the behaviour of the intact system has led the authors to conclude that the artificial device replicates the behaviour of the replaced component, which amounts to obtaining a good description of its input–output behaviour. But this inference can be safely made only if nothing has changed in the rest of the system—in other words, if the intact and the hybrid system differ from one another in the replaced component only. Indeed, plasticity could have made the lamprey adapt to a prosthesis behaving differently from the replaced component (plasticity is essential in therapeutically driven prosthetics for this very reason: this is the "blessing for engineers" discussed

by Craver). In that case, proficient stabilization abilities could therefore be explained by appealing to the lamprey's capacity to adapt to a "wrong" device, rather than to the fact that the device's behaviour actually matches the replaced component's behaviour.

Chirimuuta argues that "plastic changes are a pervasive feature of BCI research and are actually required for the correct functioning of the technology" (p. 620): it is by virtue of neural plasticity processes that nervous systems progressively adapt to "new" electro-mechanical limbs, and become able to reliably control them. Chirimuuta is right on this point: plastic changes occur in the vast majority of bionic studies reported in the literature, including Carmena et al. (2003) and all the studies mentioned in this article (with the only exception of the lamprey study). Therefore, Chirimuuta argues, all these studies violate the "no-plasticity" constraint and they should therefore be rejected as methodologically unsound on the basis of the methodological analysis carried out in (Datteri 2009). In sum, in her opinion, the no-plasticity constraint "rules out a vast swathe of BCI research as not informative in the modelling of actual biological systems for sight, hearing or reaching" (p. 622).

Yet the BCI studies mentioned above have led to interesting insights into mechanisms of motor control in primates. Therefore, if Chirimuuta is right in claiming that the "no-plasticity" constraint rules out these studies as uninformative for the modelling of actual biological systems, this is indeed bad news for the "no-plasticity" constraint and, more generally, for the entire methodological analysis proposed in (Datteri 2009), which should consequently be rejected as excessively restrictive.

However, even though it is true that plastic changes occur in the vast majority of bionic preparations, Chirimuuta is not right in claiming that all these bionic preparations violate the "no-plasticity" constraint. This is because (1) the "no-plasticity" constraint is placed on the simulation-replacement method *only*, and (2) the BCI studies she mentions (e.g., Carmena et al. 2003) are *not* instances of that methodology. For these reasons, the "no-plasticity" constraint does *not* rule out the studies in question as uninformative for neuroscientific research (in fact, it has no bearing on them whatsoever).

The fact that the "no plasticity" constraint is placed on simulation replacement studies only (point 1) follows from the discussion above. It is needed to infer a description of the behaviour of the replaced component from comparisons between hybrid system and intact system behaviours, an inference that characterizes the simulation replacement methodology only. It is intended to rule out the alternative explanations of behavioural proficiency discussed above. There are no reasons to place this constraint on other bionics-supported methodologies.[7] In the stimulation-connection methodology, in particular, one does not theorize over the behaviour of the replaced component (Sect. 3.2), and the behaviour of the target component is

[7] Datteri (2009) is explicit in placing the "no-plasticity" constraint on the simulation-replacement method only. This constraint is called "ArB2" there, ArB being the label used to refer to the simulation replacement methodology. It is one of the constraints under which "H may provide experimental support for the hypothesis that component b1 (i.e., the component *removed* from B to obtain H) behaves as MB prescribes" (p. 310, emphasis added), H, B, and MB being the hybrid system, the target biological system, and the mechanistic model under scrutiny respectively.

not inferred only from the outcome of a comparison between the behaviour of the hybrid and the intact system as in the simulation method (Sect. 3.3). No alternative explanations of the type mentioned above can be sensibly formulated in this kind of studies.

It follows from this discussion that the "no-plasticity" constraint rules out "as not informative in the modelling of actual biological systems for sight, hearing or reaching" only those studies which conform to the simulation-replacement methodology *and* in which plastic adaptation processes do occur. The lamprey study conforms to the simulation-replacement methodology, but no significant violation of the "no-plasticity" constraint is reported in (Zelenin et al. 2000). Indeed, the behaviour of the bionic system was observed right after bionic implantation, *without any training stage*—perhaps, exactly in order to exclude the occurrence of plastic changes compensating for a wrongly tuned device. Due to the absence of training, no significant plastic adaptation is likely to have taken place. The "no plasticity" constraint has not been violated there.

Does the "no-plasticity" constraint rule out a vast swathe of bionics-supported studies which have led to important discoveries on the mechanisms of brain plasticity, including (Carmena et al. 2003)? The answer crucially depends on whether these studies conform to the simulation-replacement methodology or not. As Chirimuuta (2013)'s answer to the former question is affirmative, I surmise that her answer to the latter question is affirmative too, namely that, e.g., (Carmena et al. 2003) is a simulative-replacement study. However, the argument she provides in support to this claim is that "there is no schematic difference" (p. 623) between the bionic preparation involved in the lamprey study and many other motor bionic systems. More precisely, her point is that the hybrid lamprey involved in (Zelenin et al. 2000) is structurally similar to every other bionic systems, as it is obtained by functionally replacing a biological component with an artificial device. Therefore, in Chirimuuta's view, the "no-plasticity" constraint should be placed on every bionics-supported study—with bad consequences for the constraint itself.

A reflection on Chirimuuta's argument is useful to place further emphasis on some of the claims made above. For, as pointed out in Sect. 3.1, it is true that the lamprey and the monkey preparations are "structurally similar" to each other in their being obtained by functionally replacing a component with an artificial device. But this is not sufficient to conclude that the two studies are methodologically similar to each other: indeed, as argued before, they are not. And the fact that the "no plasticity" constraint must be placed on the simulation-replacement methodology does not imply that it is to be placed on the stimulation-connection methodology too. For these reasons, her emphasis on structural similarity is neither sufficient to conclude that the "no-plasticity" constraint must be placed on a vast swathe of bionics-supported studies, nor to conclude that it rules out all these studies as "not informative in the modelling of actual biological systems". On the contrary, the arguments I have provided above provide good reasons to believe that the "no-plasticity" constraint should *not* be placed on these studies.[8]

[8] Note that the simulation replacement method has, to the best of my knowledge, only been applied *once*, namely in the lamprey study reported in (Zelenin et al. 2000). At some point in her article, Chirimuuta

Note that, in her article, Chirimuuta offers additional arguments against the "no-plasticity" constraint: in her opinion, it "unwittingly rules out numerous experimental paradigms in behavioural and system neuroscience which also elicit neural plasticity". The experimental paradigms she refers to involve in vitro, non-bionic experiments in which neural tissues are extracted and kept alive in suitable conditions. Plastic changes occurring in these tissues after some sort of training are analysed by means of electrophysiological techniques. Chirimuuta is right in pointing out that a neural tissue prepared in this way "is an experimentally modified brain, analogous to the brain that has been modified due to the introduction of a bionic implant" (p. 625). But it is not safe to infer from this premise that "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" (p. 625).

To understand, note that the non-bionic preparations she refers to are structurally similar to the systems used in the monkey and in the lamprey study. Indeed, these non-bionic preparations involve a neural tissue that is modified by *something else*, similarly to preparations in which brains are modified by *a prosthetic implant* (that is to say, in both cases one has a biological component and a modifier). But they are used in a way that is akin to the stimulation-connection methodology, to which the "no-plasticity" constraint does not apply. This is because, in the non-bionic methodologies she refers to, the focus of inquiry is on the neural tissue and not on what the modifier replaces: these methodologies are meant to inquire on what happens in the neural tissue *altered by* the modifier, exactly as in stimulation-connection studies, in which one inquires on what happens in the brain after connection to the bionic prosthesis. This is sufficient to conclude that the "no-plasticity" constraint does not apply to these non-bionic cases and that, more generally, it does not "unwittingly rules out numerous experimental paradigms in behavioural and system neuroscience which also elicit neural plasticity". What counts for the application of the "no-plasticity" constraint is methodological similarity, not only structural similarity.

5 Conclusions

Bionic systems, obtained by connecting biological tissues to artificial devices, can be used in various ways to discover biological mechanisms. In this article I have compared two bionics-supported methodologies for the study of the brain, which differ from one another in the nature of the research questions they allow one to

Footnote 8 continued
voices the suspicion that the methodological analysis offered in (Datteri 2009) has nothing to say about a vast class of bionic experiments reported in the literature. While looking outside of the mainstream can point up novel research approaches, supplementing earlier analyses and stimulating methodological discussion of novel and emerging fields previously overlooked by philosophers, I agree with her on the fact that (Datteri 2009) covers a minimal part of the contemporary bionics-supported research. The aim of this article is to examine the structure of a methodology which is much more often adopted in contemporary scientific research than the simulation-replacement one.

address and in the experimental procedure (see Table 1). I have also argued that not every bionics-supported study makes an epistemic use of prostheses *qua* functional replacers of biological components. These distinctions have been brought to bear on some claims on the epistemic value of bionics made in the philosophical literature.

The analysis proposed here can be refined by addressing a number of epistemological and methodological questions concerning stimulation-connection studies which are not discussed here. What auxiliary assumptions are needed to infer theoretical conclusions on the non-replaced part of a biological system from the result of bionic experiments conforming to this methodology? What criteria guide inferences from the analysis of plastic changes occurring in the brain after connection with a *robotic* device to the theoretical modelling of plastic changes occurring in the brain during control of a *biological* limb? And, more generally, what kind of theoretical questions on the non-replaced part of the hybrid system can be fruitfully addressed through this methodology? I have claimed that bionic experiments can assist in modelling the input–output behaviour of biological components and their internal mechanisms. Chirimuuta (2013) has argued that bionic systems can distinctively assist in the discovery of *organizational principles* rather than of mechanistic models. In her view, organizational principles do not concern "the layout of an actual neural circuit or mechanism", but rather "the operational principles that allow a range of neuronal mechanisms to do what they do" (p. 629). I agree with these claims. However, as pointed out before, the analysis of plastic changes in the monkey study as well as, for example, the analysis of the behaviour of neurons not directly participating in motor control in other studies discussed here, has led to the formulation of conjectures on the layout of the sensory-motor mechanisms implemented in the brain. Over and above these particular remarks, identifying and analysing bionics-based experimental procedures, reasoning on how empirical data are brought to bear on the hypotheses under scrutiny, and isolating the methodological and epistemological auxiliary assumptions needed to draw theoretical conclusions from the analysis of the behaviour of bionic systems, may contribute to the piecewise understanding of the various ways these new and emerging technologies can contribute to basic neuroscientific research.

References

Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., et al. (2003). Learning to control a brain–machine interface for reaching and grasping by primates. *PLoS Biology, 1*(2), 193–208.

Chapin, J. K., Moxon, K. A., Markowitz, R. S., & Nicolelis, M. A. L. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neuroscience, 2*(7), 664–670.

Chirimuuta, M. (2013). Extending, changing, and explaining the brain. *Biology and Philosophy, 28*(4), 613–638.

Chou, P. C., & Hannaford, B. (1997). Study of human forearm posture maintenance with a physiologically based robotic arm and spinal level neural controller. *Biological Cybernetics, 76*(4), 285–298.

Cordeschi, R. (2002). *The discovery of the artificial behavior mind and machines before and beyond cybernetics*. Dordrecht: Springer.

Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press.

Craver, C. F. (2010). Prosthetic Models. *Philosophy of Science, 77*(5), 840–851.

Datteri, E. (2009). Simulation experiments in bionics: a regulative methodological perspective. *Biology and Philosophy, 24*(3), 301–324.

Datteri, E., & Tamburrini, G. (2007). Biorobotic Experiments for the Discovery of Biological Mechanisms. *Philosophy of Science, 74*(3), 409–430.

Floreano, D., Ijspeert, A. J., & Schaal, S. (2014). Robotics and Neuroscience. *Current Biology, 24*(18), R910–R920.

Ganguly, K., & Carmena, J. M. (2009). Emergence of a stable cortical map for neuroprosthetic control. *PLoS Biology, 7*(7), e1000153.

Ganguly, K., Dimitrov, D. F., Wallis, J. D., & Carmena, J. M. (2011). Reversible large-scale modification of cortical networks during neuroprosthetic control. *Nature Neuroscience, 14*(5), 662–667.

Glennan, S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 443–464.

Golub, M. D., Chase, S. M., Batista, A. P., & Yu, B. M. (2016). Brain–computer interfaces for dissecting cognitive processes underlying sensorimotor control. *Current Opinion in Neurobiology, 37*, 53–58.

Grasso, F. W., Consi, T. R., Mountain, D. C., & Atema, J. (2000). Biomimetic robot lobster performs chemo-orientation in turbulence using a pair of spatially separated sensors: Progress and challenges. *Robotics and Autonomous Systems, 30*(1–2), 115–131.

Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., et al. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature, 442*(7099), 164–171.

Koralek, A. C., Jin, X., Long, J. D., II, Costa, R. M., & Carmena, J. M. (2012). Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. *Nature, 483*(7389), 331–335.

Le Masson, G., Renaud-Le Masson, S., Debay, D., & Bal, T. (2002). Feedback inhibition controls spike transfer in hybrid thalamic circuits. *Nature, 417*(6891), 854–858.

Lebedev, M. A., Carmena, J. M., O'Doherty, J. E., Zacksenhouse, M., Henriquez, C. S., Principe, J. C., et al. (2005). Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain–machine interface. *The Journal of Neuroscience, 25*(19), 4681–4693.

Newell, A., & Simon, H. A. (1961). Computer Simulation of Human Thinking: A theory of problem solving expressed as a computer program permits simulation of thinking processes. *Science, 134*(3495), 2011–2017.

Nicolelis, M. A. L. (2003). Brain–machine interfaces to restore motor function and probe neural circuits. *Nature Reviews Neuroscience, 4*(5), 417–422.

Nicolelis, M. (2011). *Beyond Boundaries: The New Neuroscience of Connecting Brains With Machines And How It Will Change Our Lives*. New York: Times Books.

Orsborn, A. L., & Carmena, J. M. (2013). Creating new functional circuits for action via brain–machine interfaces. *Frontiers in Computational Neuroscience, 7*, 157.

Rosenblueth, A., & Wiener, N. (1945). The Role of Models in Science. *Philosophy of Science, 12*(4), 316–321.

Simon, H. A., & Newell, A. (1962). Computer Simulation of Human Thinking and Problem Solving. *Monographs of the Society for Research in Child Development, 27*(2), 137–150.

Suppe, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. Urbana and Chicago: University of Illinois Press.

Tamburrini, G., & Datteri, E. (2005). Machine Experiments and Theoretical Modelling: from Cybernetic Methodology to Neuro-Robotics. *Minds and Machines, 15*(3–4), 335–358.

Wander, J. D., & Rao, R. P. N. (2014). Brain–computer interfaces: A powerful tool for scientific inquiry. *Current Opinion in Neurobiology, 25*, 70–75.

Webb, B. (2001). Can robots make good models of biological behaviour? *The Behavioral and Brain Sciences, 24*(6), 1033–1050.

Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., et al. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature, 408*(6810), 361–365.

Woodward, J. (2002). What Is a Mechanism? A Counterfactual Account. *Philosophy of Science, 69*, S366–S377.

Zelenin, P. V., Deliagina, T. G., Grillner, S., & Orlovsky, G. N. (2000). Postural control in the lamprey: A study with a neuro-mechanical model. *Journal of Neurophysiology, 84*(6), 2880–2887.