Studies in History and Philosophy of Biological and Biomedical Sciences 53 (2015) 94–101

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Can functionality in evolving networks be explained reductively?

Ulrich Krohs

Department of Philosophy, Center for Philosophy of Science, University of Münster, Domplatz 6, 48143 Münster, Germany

ARTICLE INFO

Article history:
Available online 17 April 2015

Keywords:
Design principle
Function
Multiplicity
Irreducibility
Gene regulatory network
Neural network

ABSTRACT

Philosophers of biology disagree about an adequate explication of the concept of function. Instead of perpetuating the debate on the level of in principle-arguments, this paper aims first at reconstructing functional talk in the biological research papers of Marom and Braun, which focus on two different kinds of evolving networks, and in discussing the ontological consequences which the authors draw from their results. Marom investigates evolving neural networks controlling Braitenberg vehicles. Braun observes the evolutionary rearrangement or "rewiring" of the genetic network of genetically modified yeast on a short time scale. In both cases, the parameters under investigation are defined in functional terms. However, both authors report striking differences in the structures that realize one and the same function, as well as striking differences in the function of identical structures. From this, they construct an argument against reductionism. The second aim of my paper is an inquiry into the epistemic legitimacy of this conclusion. This requires addressing critically several concepts on which Marom and Braun's argument is built.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Biological and Biomedical Sciences*

1. Introduction

Braun and Marom (in this issue) interpret two experimental results as challenging reductionism. One is a reported lack of separation of time scales on different levels of organization of a biological system, which goes hand in hand with a lack of separation of levels of the interacting entities themselves. The other is what they call (deep) two-way microscopic–macroscopic degeneracy: any functional structure on the micro-level may serve or contribute to various higher-level functions, and each higher-level function may be realized by various micro-structures. As Bechtel (in this issue) shows, scale freeness does not prohibit stating mechanisms as idealized structures that bring about a certain behavior of the network. Even a scale-free network can be structured according to pragmatic or instrumentalist criteria in order to explain its behavior. If nature turns out to be continuous in a certain respect, we may nevertheless describe it as partly discrete in order to explain what is going on. I will not be concerned with this issue in this paper. The focus is on functionality and on the indeterminacy or ambiguity of the mapping of functions on their microscopic realizers as described by Braun and Marom, and on the connection of indeterminacy to the normativity of function talk.

Braun and Marom claim that their observation of the ambiguity of the mapping of functions to function bearers challenge reductionism. I will redescribe their results in the well-established terminology of multiple realizability and heterogeneity of functions (e.g., Carrier, 2000). This will show that there is nothing unusual with Braun and Marom's "two-way microscopic–macroscopic degeneracy" so that any irreducibility claim falls within the scope of the classical debate of these issues, as long as it is not developed into a view of "anomalous functionality" which prevents *any* generalization of the observed effects.

After discussing reference to functionality and design in the papers of Marom (Section 2) and Brown (Section 3), I offer an account of function that links the concept of function to the concept of design in a way that is compatible with the biologists's writings (Section 4). I then discuss the issue of the stated microscopic–macroscopic degeneracy and its relevance for the question of

E-mail address: ulrich.krohs@uni-muenster.de.

http://dx.doi.org/10.1016/j.shpsc.2015.03.009

reductionism (Section 5). This leads in the conclusion (Section 6) to modesty in drawing reductionist or anti-reductionist conclusions from the experiments discussed.

2. Reference to functionality in Marom's research papers on neural networks

2.1. Design principles

In a highly interesting series of papers, Marom and his colleagues have developed a system in which individualized biological neurons form a self-organized network, which controls, *via* electronic circuits, a mechanically realized Braitenberg vehicle[1] (see Marom & Shahaf, 2002; Shahaf et al., 2008). To control the vehicle is thus the function of the network, which it acquires in a learning process. Marom denies that the neural network[2] can be analyzed in a way that fully explains how it really works (Marom et al., 2009). The argument is based on the example of a simplified case in which no learning is involved. Marom's argument against the possibility of reverse engineering runs in two steps. In a first step, he states that there might be no way to infer the rules that govern the behavior of the network if the class of possible rules is too large, and that this holds even after the observation of an infinite number of instances. (I take it that *inference* is to be understood as *inference to the best explanation*, since the claim would otherwise run the risk of being trivially true.) In the second step, Marom shows that experimental data on the neuronal activities in the network may support an inference to an explanation that is actually wrong. The case is nicely construed: the authors use a vehicle in which the actual functional dependency of the electronic part of the controlling device on the output of the network of neurons is known. Analysis of the network by methods of reverse engineering, however, supports the inference to a different type of coupling: It supports the inference to alleged design principles which were in fact not implemented by the experimenters, rather than to those that were implemented.

Marom points out that, in this case, a valid inference leads to a wrong conclusion.[3] Moreover, he concludes from this case of "wrong induction" (Marom et al., 2009) that it is impossible to determine correctly design principles from complex biological networks by means of reverse engineering. The question thus is whether the asserted failure to infer the correct design principles is indeed fatal for the project to reverse-engineer complex biological networks.

Green (in this issue) argues that asking for design principles with respect to complex biological networks might mistake the dynamic structure of those networks anyway. To her, failing to recover the "actual" design principles may be the adequate result of certain attempts of reverse engineering rather than a proof that reverse engineering methods are inapplicable in such cases. I fully agree with her argument. But I see a more fundamental problem with Marom's argument, due to his equivocal use of "design principle".

When Marom claims that the case study shows that we cannot infer the right design principles from the experiments, he uses the term "design principle" in two different ways. Required is, thus, a disambiguation of the relevant terms. Initially, "design principle" denotes the principles that the engineers wanted to implement in the control unit of the Braitenberg vehicle. So "design" here refers to the *process* of designing—or of developing—a complex entity. However, when analyzing the behavior of the system, one is confronted with the fully setup system only rather than with its genesis or etiology. "Design" now refers to the *actual* structure and functioning of the complex system, which is the *result* of the designing and construction process. Both concepts must be discerned thoroughly (Krohs, 2009). It is by no means clear that those principles which governed designing are fully present in the designed entity, nor that no other principles emerge in the actual design of the system which were not explicitly put into it by planning. If we discern *design* principles in the strict sense from *designing* principles, the situation looks much less threatening for reverse engineering. It may indeed be impossible inferring the principles that governed *designing* by analyzing merely the principles that govern the behavior of the product of this process, but this does in no way show that researchers draw any false conclusions on the *design*. In the example, the imagined investigators are justified in inferring *design* principles, though they are not in a position to infer *designing* principles.

One might ask, however, whether the actual design principles in the example might not be provably identical with the designing principles. I dare doubting this. One reason is that the system is opaque to its designer. He/she may have fixed the designing principles, but he need not know the actual design principles. The system may run, for many reasons, different from his/her intentions, which is the rationale behind the wise advice: "never touch a running system!" The system may also have two equivalent descriptions, both of which being adequate. (Compare this to Marom's claim about the "deep degeneracy" of functionality, which will be discussed below.)

If, on the other hand, a description proceeds on a design hypothesis which does not accord with the *designing* principles *and* fails to describe the actual behavior of the system, I do not see any reason why this could not be discovered by further thorough empirical research. If the design of a system is a target of empirical investigation at all, which Marom presupposes and with which I agree, then any mistaken inference means first of all that the hypothesis was not tested hard enough. Presenting an experimental setting in which the actual results justify the inference to a flawed explanation cannot prove a claim about the unrecognizability of design principles. This required a proof that all possible experimental results, or at least all results of reasonably feasible experiments, would support the wrong inference, which is not to be expected in the present case.

2.2. Concepts of function

In Marom et al. (2009), among the co-authors of which is also Braun, the neuro-system is called "functional." Since an extended debate among philosophers of biology is devoted to the concept of function, taking notice of the actual use that biologists make of it is most welcome. The spectrum on offer ranges from teleological concepts like the etiological ones (Millikan, 1984; Neander, 1991; Wright, 1973) to the (seemingly innocent) concept of a function as a role that some component fulfills in a system. In the latter case, function ascription might even be considered in a purely

[1] The setting uses Braitenberg vehicles II, i.e., vehicles with separately regulated left and right motors and only activating influences of sensors to motors (Marom et al., 2009).

[2] When talking about neural networks, I am referring to such networks of real biological neurons of organismic origin throughout this paper, not to computer simulations or to mathematical structures.

[3] To give the reader an idea about the kind of design principles at stake: the system is designed to react to the rank order of the first spikes of a predefined subset of neurons. But unfortunately, if you test the alternative hypothesis that the response depends on the population response rate, the data on neuron activity strongly suggest that this hypothesis should be accepted, i.e., that the population response rate rather than the rank order of the first spikes triggers the behavior of the vehicle. Further details of the experiments, which do not matter for my argument, are reported in Green (in this issue).