UNIVERSITAT DE BARCELONA

Methodological strategies in contemporary symbiosis research and their historical roots: From mechanistic to non-mechanistic modes of explanation

Javier Suárez Díaz

Methodological strategies in contemporary symbiosis research and their historical roots: From mechanistic to non-mechanistic modes of explanation

Estrategias metodológicas en la investigación contemporánea sobre la simbiosis y sus raíces históricas: De las explicaciones mecanicistas a las explicaciones no mecanicistas

Doctoral thesis submitted by **Javier Suárez Díaz** to the University of Barcelona for the degree of Doctor of Philosophy

Supervisors: José A. Díez Calzada and John A. Dupré

PhD Program: Ciència Cognitiva i Llenguatge
Research Line: Filosofìa del Llenguatge, de la Mente i de las Ciències Cognitives
Facultat de Filosofia
Universitat de Barcelona
January 2020

RESUMEN

La investigación sobre la simbiosis es un campo en crecimiento en la biología contemporánea. Los avances actuales en el modelado y las técnicas experimentales han permitido desarrollar nuevas maneras de estudiar algunos sistemas simbióticos de múltiples especies cuyo estudio había sido ignorado en el pasado. Algunas de estas nuevas técnicas experimentales y de modelado se basan en el uso de herramientas matemáticas sofisticadas (como el análisis de redes) que solo se pueden emplear si el sistema se concibe de manera holística. Una de las principales consecuencias de este enfoque es que la descomposición del sistema en partes se vuelve una tarea imposible: si los biólogos quieren comprender cómo funcionan algunos sistemas simbióticos de múltiples especies, cómo se comportan o cómo evolucionan, deben estudiar el sistema de manera holística, en lugar de estudiar cómo cada una de las partes del sistema se relaciona entre sí. Este tipo de investigación parece cuestionar la tradición dominante en la filosofía contemporánea de la ciencia y la biología, a saber: el *nuevo mecanismo*. De acuerdo con los principios de la filosofía del nuevo mecanicismo, los sistemas biológicos deben descomponerse en sus elementos componentes para comprender adecuadamente cómo funcionan y para descubrir las conexiones causales entre los mismos. El propósito de esta tesis doctoral es comprender cómo la investigación contemporánea sobre simbiosis cuestiona algunas de las tesis filosóficas centrales que subyacen a la filosofía del nuevo mecanismo. Para hacerlo, la tesis se basará en dos métodos filosóficos: (1) análisis de la práctica científica; (2) análisis conceptual. Al aplicar estos dos métodos a la investigación contemporánea sobre simbiosis, la tesis da lugar a tres artículos en revistas especializadas, que se agregan como anexos. La contribución original clave de la disertación doctoral es que la investigación contemporánea sobre la simbiosis se basa en el uso de ciertos métodos matemáticos que solo son aplicables si el sistema se estudia de manera holística y, por lo tanto, conlleva una forma de explicación no mecanicista y no causal. Al final, se presentan varias preguntas abiertas para futuras investigaciones.

ABSTRACT

Symbiosis research is a growing field in contemporary biology. Current advances in modelling and experimental techniques have made possible to develop new ways of studying some multispecies symbiotic systems whose study had been mostly ignored in the past. Some of these new modelling and experimental techniques rely on the use of

sophisticated mathematical tools (such as network analysis) that can only be used if the system is conceived holistically. One of the main consequences of this approach is that de-composition becomes an impossible task: if biologists want to understand how some multispecies symbiotic systems work, how they behave, or how they evolve, they need to study the system holistically, rather than study how each of the parts of the system relates to each other. This type of research seems to question the dominant tradition in contemporary philosophy of science and biology, namely: new-mechanism. According to the principles of new-mechanistic philosophy, biological systems need to be de-composed in their component elements to properly understand how they work, and in order to unveil the causal connections among the components. The purpose of this doctoral thesis is to understand how contemporary symbiosis research questions some of the core philosophical thesis that underlie new-mechanistic philosophy. To do so, the thesis will rely on two philosophical methods: (1) analysis of scientific practise; (2) conceptual analysis. By applying these two methods to contemporary symbiosis research, the thesis gives rise to three papers in specialized journals, added as annexes. The key original contribution of the doctoral dissertation is that contemporary symbiosis research relies on the use of certain mathematical methods that are only applicable if the system is studied holistically, and thus entail a form of non-causal-mechanistic explanation. In the end, several open questions for future research are presented.

Keywords

Symbiosis research · Scientific explanation · Scientific methodology · Philosophy of Science · Philosophy of Biology · New-mechanism · Philosophy in Science

The doctoral project has been funded by the Ministry of Education, Government of Spain, under the FPU Schema (grant agreement FPU16/02570), and is part of the project ‘Laws, Explanation and Realism in the Physical and Biomedical Sciences’, funded by the Ministry of Economy and Competitiveness, Government of Spain (grant agreement FFI2016-79799-P).

I. OVERVIEW OF THE DOCTORAL PROJECT

The aim of this OVERVIEW is to contextualize the doctoral project, carefully reviewing the scientific problem that the project studies and motivating it in the context of contemporary philosophy of science. At the same time, this overview will provide an introduction to the research goals of the dissertation, the degree of success in reaching them, and the type of open questions that have emerged after it. Finally, it will also explain the scientific methodology that has been employed and justify its use for investigating the research questions that gave rise to the project.

The structure will be as follows. In section 1, I will introduce the scientific problem to be studied (section 1.1), the philosophical problem that will be investigated (section 1.2), I will indicate how the scientific problem and the philosophical problem are related (section 1.3), and I will highlight the fields that the current project will affect (section 1.4). In section 2, I will explain the research goals of the doctoral project, and the guiding hypothesis that have been used to frame the investigation. In section 3, I will explain –and justify– the research methodology that has been applied. In section 4, I will present the papers, summarize their main discoveries, and explain how they fulfil the research goals and conform to the guiding assumptions of the project. Finally, in section 5, I present the open questions that the project has given raise to, and that motivate future research.

1. Introduction to the research problem investigated in the doctoral thesis

1.1. Description of the scientific problem: How symbiosis research has led to the development of holistic methodologies in the study of biological systems

In the last decades, due to the substantial advances in the study of microorganismal life on Earth, the study of its influence on animal and plant biology, and the pervasiveness of symbiosis, scientists have developed new methodological strategies to study the biology of these multispecies biological systems, including aspects such as their behaviour and evolutionary dynamics (Gilbert et al. 2012, 2017; Stencel & Proszewska 2017; Bapteste & Huneman 2018). These symbiotic systems have been referred to as ‘holobionts’, when they consist in an animal/plant host plus its symbiotic microbiome, and ‘microbiomes’, when they consist in a collection of microorganisms living in a concrete ecological niche. They constitute a genuine biological reality that challenges us to revisit some of the most well-established work in genetics –prompting e.g. the appearance of *community genetics*– , as well as some important philosophical concepts such as the notion of ‘unit of

selection', 'biological lineage', or the concept of 'biological individuality' (Dupré & O'Malley 2009; Clarke 2010; Pradeu 2012; DiFrisco 2017; Lidgart & Nyhart 2018; Gissis et al. 2018).

The contemporary increase in our knowledge of the microorganismal world, and especially of its importance for sustaining plant and animal life on Earth as we know it –including several aspects of human health– has led a world-leading group of researchers to coin the 'hologenome concept of evolution' (Zilber-Rosenberg & Rosenberg 2008; Rosenberg & Zilber-Rosenberg 2014; Theis et al. 2016). The hologenome concept of evolution postulates that the *holobiont*, the biological symbiotic entity composed of the animal or plant host, plus the set of microbes that compose its microbiome, is a biological individual *and* a unit of selection in evolution (Lloyd 2017; Roughgarden et al. 2018; Lloyd & Wade 2019; Suárez & Triviño 2019). Grounded on this hypothesis, defenders of the hologenome concept are committed to the claim that the evolution of animals and plants is directly connected to the evolution of their microbiomes, such that many of the traits that allow animal/plant specializations to specific ecological niches are encoded in their microbiomes, rather than in the animal/plant genome –for interesting examples see Mendoza et al. (2018), and Gilbert (2019), Chiu & Gilbert (2019).

The hologenome hypothesis has led biologists to develop *new research techniques*, *new modelling methods*, and *new experimental frameworks* to test the validity of the hologenome concept, oriented to discover whether (and, if so, *how*) the microbiome has affected animal and plant evolution (Sharon et al. 2010; Brucker & Bordenstein 2013; Mendoza et al. 2018; Bapteste & Huneman 2018). In addition to the evolutionary research undertaken in agreement with the hologenome framework, as holobionts and microbiomes are multispecies *communities*, some current research on the microbiome has been oriented *ecologically* (rather than evolutionarily), with the aim of studying the ecological properties of the microbiome (e.g. how bacterial species interact, how they affect each other, what's their degree of stability and it is maintained, etc.) (Coyte et al. 2015; Foster et al. 2017).

Irrespectively of whether these systems are studied evolutionarily or ecologically, it is clear that empirical advances in hologenome and microbiome research have led biologists to recognize the importance of *holistic perspectives* in biology. That is, these discoveries have prompted biologists to realize that the properties of certain biological systems *need* to be studied holistically –or systemically– rather than part-by-part through a process of decomposition. Holistic methodologies have become pervasive in the study of biological systems, especially in the study of the well-known phenomenon of symbiosis. This contemporary development in scientific practice is far from trivial for the development of contemporary philosophy of science, for it questions some of the basic assumptions of the well-established tradition of *new-mechanism*, which has been the dominant tradition in some branches of philosophy of science since early 2000, when "Thinking about mechanisms" was first published (Machamer et al. 2000).

Giving this context, this doctoral dissertation aims to provide an answer to the following question, which guided the research activities: How will the development of contemporary symbiosis research –with special emphasis on the *methods* that are employed to investigate symbiosis– affect philosophy of science?

1.2. Description of the philosophical problem: Why holistic methodologies matter for the development of contemporary philosophy of science. A contrast with the new-mechanistic tradition

Since 2000, when Machamer, Darden and Craver published their famous paper "Thinking about mechanisms" (Machamer et al. 2000), the so-called new-mechanistic philosophy has been the standard research tradition in the study of scientific explanation within philosophy of science and biology.[1] According to the latter, scientific research must be guided by a mechanistic paradigm, and to explain a phenomenon consists in uncovering the mechanisms that *causally* produce it. New-mechanism has evolved a lot since the first new-mechanistic hypothesis was formulated, and different ideas about the notion of "mechanism", its relevance in contemporary philosophy of science, the limits of its applicability, etc. have been formulated (Glennan 2002; Bechtel & Abrahamsen 2005; Craver 2007; Craver & Darden 2013; Glennan & Illari 2017; Krickel 2018). Despite the differences between distinct new-mechanistic accounts, it can be argued that all new-mechanists share three key principles that define their general philosophical commitments (see Nicholson 2012, 2013; Militello & Moreno 2018; Huneman 2018):

> (C1) A commitment to *a causal view of the world*
>
> (C2) A global *rejection of the Hempelian idea that scientific explanations take the form of logical arguments* that can proceed either inductively or deductively
>
> (C3) The assumption that *mechanisms provide the causal ingredient* that is required to make every scientific explanation/hypothesis philosophically acceptable.

The combination of these three commitments, together with a specific concept of "mechanism", provide the basic resources to understand new-mechanistic philosophy, as well as its relevance in contemporary science, philosophy of science *and* general philosophy. For new-mechanism is not only a theory about how scientific explanations must proceed, or how scientific research must be carried out. It is also a global theory about some of the key philosophical concepts, such as "causality" [(C1) & (C3)].

Let us start from the beginning. New-mechanist philosophers believe that proper scientific explanations must proceed by discovering the mechanism (or set of mechanisms, if more than one) that *causally* produce the phenomenon of investigation [(C1)]. Now, an important question arises: what is a mechanism? According to new-mechanism, a mechanism is a structure that consists in a series of entities (or parts) and activities (or operation) organised in a way such that the interactions between the entities causally produce a set of regular changes that bring the structure from a set of initial conditions to a termination condition (Machamer et al. 2000; Glennan 2002; Bechtel & Abrahamsen 2005; Craver 2007; Craver & Darden 2013; Glennan & Illari 2017; Krickel

[1] There are at least two clear earlier hypotheses about the importance of mechanisms in the biological sciences: Bechtel & Abrahamsen (1993), and Glennan (1996). Nonetheless, as these works acquired special relevance after the publication of "Thinking about mechanisms", I will assume that the new-mechanist tradition starts then.

2018; Deulofeu & Suárez 2018).[2] This definition includes three elements: a) entities, or parts, that interact in the mechanism and that must be singled out in the scientific explanation; b) activities, or operations, which refer to the way in which the entities interact with each other; c) organization, which refers to the particular manner how the entities and the activities are spatiotemporally distributed in the mechanism. In addition to these three elements, a mechanism must consist in a structure such that the causal interactions between the parts result in in the system moving from a set of initial or set-up conditions, to a set of finish or termination conditions (**Figure 1**).

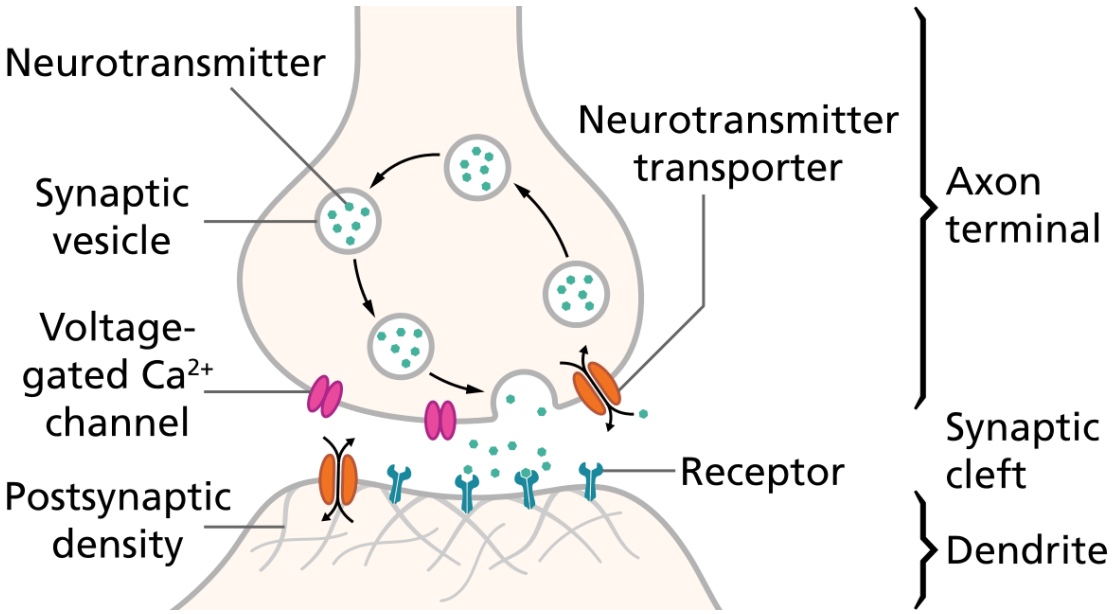

Figure 1. Canonical example of a mechanism –the neural synapsis. The parts (or entities) and their operations (or activities) are particularly salient in the picture. At the same time, the organisation of the system in three distinguishable elements (the axon terminal, the synaptic cleft, and the dendrite) is salient, and serves to clearly explain how the mechanism works (i.e. how the electric signal has to travel from the axon terminal to the dendrite, and this needs to be chemically mediated through the synaptic cleft). Based on the example in Machamer et al. (2000: Figure 1). Image taken from Wikipedia Commons, open license.

This view of mechanisms is far from philosophically and scientifically irrelevant, for it relies on a very specific view of the scientific process (with a very important normative dimension too, see e.g. Kaplan & Craver 2011). Concretely, new mechanist philosophers share the following methodological commitment:

[2] There is currently a controversy between new-mechanistic philosophers about how different philosophers conceive the concept of mechanism, the type of commitments of new-mechanistic philosophy, etc. Even though this may be important for certain aspects of the development of the doctoral project (see especially "Paper 2"), it is irrelevant at this point, for I only aim at introducing a general view of the commitments of new-mechanistic philosophy, and the type of research they (*all*) build upon. For contemporary research on how different mechanistic philosophers define the concept see Nicholson (2012, 2013, 2018), Glennan & Illari (2018)

> (C4) The belief that the analytic decomposition of a complex system into their simpler components permits that relatively few parts could be *methodologically isolated from the rest such that causal mappings between specific functional operations and their distinguishable structural components could be drawn.*

Let me explain carefully what (C4) means (and what it *does not* mean), for it is essential to understand the philosophical core of new-mechanistic philosophy, and thus the type of beliefs about causality, explanation, methodology, etc. that follow. For new-mechanistic philosophers, a key element in scientific research is the capacity to un-pack, or *de-compose* the system that they are studying in different parts (or entities) whose specific behaviour can be studied in isolation. In other words, according to new-mechanistic philosophy, the mechanism is a structure that performs a function –i.e. it systematically leads to a termination condition– because each of its parts engage in a set of activities in virtue of the way that these parts are organised. Thus, de-composition, or the capacity of breaking up a mechanism into its component parts is an essential element of new-mechanistic philosophy (such that it has led some authors to argue that this should be the main aim of science, see Craver 2007).

This observation about the key commitments of new-mechanistic philosophy have led Issad & Malaterre (2015: 270) to argue that a mechanistic explanation (or methodology) would work if and only if two conditions are met: (1) a *model of the mechanism* (entities, activities, organisation) needs to be provided; (2) a *causal story* of how the parts interact to produce the termination conditions needs to told. These conditions are individually necessary and jointly sufficient for a proper mechanistic explanation of the phenomena. New-mechanists believe that most scientific activity will be guided by searching mechanisms, and thus the aim of science would be to find de-composable systems, to understand how each of the parts of the system causally relates to the rest of the parts, and to explain how these interactions produce the final or termination conditions that require explanation.

New-mechanism, through its insistence on de-composition and the ability to understand how each part of the mechanism interacts with each other, conceives science as an activity that requires finding *bottom-up causes* that allow the mapping between the functional operations and the components that bring these operations about. This view, however, does not necessarily entail a form of physical reductionism à la Salmon (Salmon 1984) or an ontic interpretation of mechanistic philosophy *à la* Craver (Craver 2014). On the contrary, most new-mechanists accept that the world has a hierarchical structure, and that different mechanisms can be found at different levels, realized by different entities, activities, and their organisation (Krieckel 2018). This point is important, for otherwise new-mechanism could be easily discredited by arguing that the special sciences generate explanations that either are not ontic, or do not de-compose the system into molecular components (Deulofeu & Suárez 2018). But since the hierarchical view of the world that most new-mechanists accept does not make them fall prey of that criticism, then a solid criticism to their postulates would require finding examples of explanations at some hierarchical level that simply do not satisfy at least one of the requirements imposed by the new-mechanistic philosophy.

To summarize, new-mechanism is a view of philosophy, and of the scientific activity that is structured around four basic commitments (C1)-(C4), and that takes the search of mechanisms –understood as a set of entities, activities, and their organisation that interact in a way that be causally specified– as the basic task of scientific research. Now the question I need to ask is how and why the specific scientific problem I aim to address in this doctoral project (symbiosis research) relates to the new-mechanistic philosophy. In other words, to what extend do I believe that symbiosis research questions the new-mechanistic paradigm, and why should this be relevant in contemporary philosophy of science?

1.3. From symbiosis research to philosophy research: How contemporary symbiosis research affects the development of philosophy of science

The new-mechanistic paradigm in philosophy of science has been mainly inspired by contemporary developments in molecular biology and neuroscience. Indeed, it is not uncommon to read that the main task of molecular biology is to uncover the mechanisms that underlie a particular process. For instance, biologists have repeatedly reported to have discovered the (molecular) mechanisms of tumorigenesis (Batlle & Wilkinson 2012), or the (molecular) mechanisms of antibiotic resistance (Blair et al. 2014). And neuroscientists often talk about the mechanisms of neuronal cell death (Yuan et al. 2003), or the mechanisms of neurodegeneration in Alzheimer's disease (Crews & Masliah 2010). New-mechanists have been especially impressed by this kind of developments, which made them realise that *mechanisms* must play a central role in contemporary science. The whole project of new-mechanistic philosophy has thus become to invigorate the concept of "mechanism" and use it as a guiding notion in contemporary philosophy of science that substituted the "obsolete" concepts inherited from neo-positivism.[3]

My doctoral thesis takes symbiosis research as a fundamental field to prove the universal validity of new-mechanistic philosophy. Contemporary symbiosis research still relies on many assumptions deriving from molecular biology, and thus it seems that the new-mechanistic principles would still apply there. However, at the same time, symbiosis research expands substantially the field, and it seems that some of the key postulates of new-mechanistic philosophy will be put under question if we take seriously some of the new modelling practises that this type of research has developed (and is still developing).

Contemporary symbiosis research can be basically divided according to the type of systems that is studies: two-species systems (e.g. host-parasite interactions), and multispecies systems. The former, whose study seems methodologically closer to the new-mechanistic paradigm, consists in studying how two species can interact closely during their lifetimes. The latter, however, consists in studying how living communities composed by thousands of species (e.g. the microbiome) interact, what are their biological properties, how do they evolve, etc.

Two-species systems are sometimes studied by decomposing the system and unveiling the type of systematic connections between the interacting organisms (Bourrat

[3] New-mechanism appears in a moment in which most philosophers of science had questioned the usefulness of the concept of "scientific law" for contemporary science, and for philosophy of science (Cartwright 1983; Giere 1999; cf. Mitchell 2003).

2018; Lynch et al. 2019). However, as I will analyse in my thesis, some contemporary symbiosis research carried out on two-species systems also relies on some global properties of these systems whose study seems to be radically different to the type of methodology that new-mechanism requires. Concretely, some of these systems are studied *mathematically*, by relying on game theory and the concept of evolutionary stable strategy (Blaser & Kirchner 2007). No mechanistic methodologies seem to be applied to the study of these systems, and it becomes necessary to understand why and how this can be so.

Multispecies symbiosis systems, on the other hand, seem not to be ever studied mechanistically, or relying on a mechanistic methodology (Green & Jones 2016; Huneman 2018; O'Malley & Skillings 2018). Multispecies symbiosis systems, given their internal complexity, require different types of methodologies to be studied. Concretely, recent research on these systems is based on the use of graph theory (topological reasoning) in the form of network analysis (**Figure 2**). A key feature of the use of these techniques is that they can only be applied if the system is studied *holistically*, for the type of properties that these techniques allow to study only appear at the systemic level. In other words, the nature of symbiotic multispecies systems makes impossible to de-compose them for their study, which runs contrary to the principles of new-mechanistic philosophy.

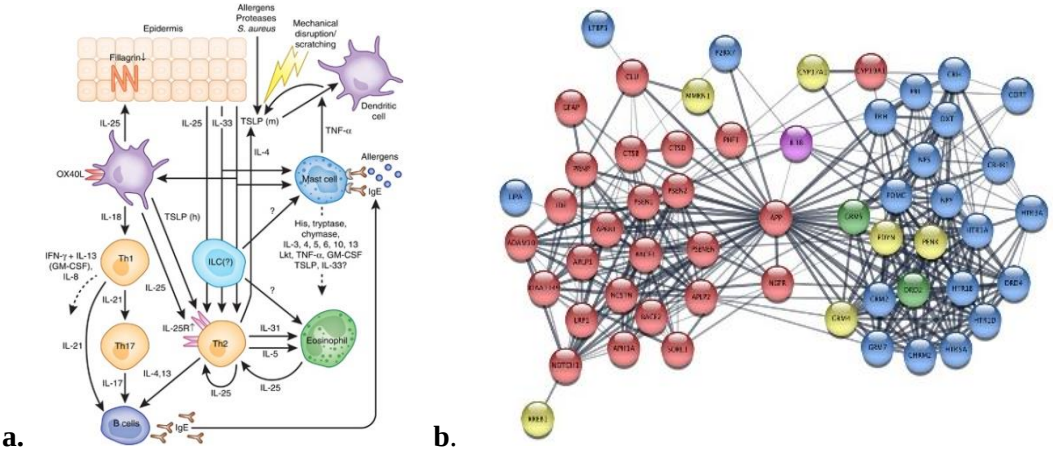

Figure 2. Comparison of a mechanistic model from molecular biology (**a**) with a network model from multispecies symbiotic research (**b**). In mechanistic models, the parts are well defined (molecules, cells, etc), and the system is de-composable (each entity plays a specific causal role that can be studied independently of the system). In a network model, however, the parts are abstract objects and the interactions realize a web-like structure that generates collective global properties that are not de-composable, i.e. the causal role of each part cannot be studied separately from the system, for each part only behaves the way it does when it is embedded within the system. Specific details about the biological nature of the systems are irrelevant. The images have been taken from https://www.slideshare.net/larsjuhljensen/network-visualization-a-crash-course-on-using-cytoscape.

A second, and very important point of multispecies symbiosis research is that it does not rely exclusively on the construction of bottom-up methodologies. On the contrary, multispecies symbiosis research relies many times on top-down methodological

approaches. The difference between the two types of approaches is as follows. Bottom-up approaches consist in de-composing the system into its basic molecular constituents to understand how these can cause/produce the type of macroscopic behaviour that we observe. For instance, investigating how a genetic make up may cause a disposition to develop a disease is a bottom-up approach, in which the molecular components are attributed causal efficiency in producing the macroscopic effect. Multispecies symbiosis research, however, relies in many cases in top-down approaches to argue that these macroscopic effects are indeed causally responsible of the molecular patterns that we observed. For instance, it is usually argued that a concrete macroscopic pattern (like inflammatory bowel disease) is responsible of the strange distribution of species in the microbiome, and not the other way around. Of course, top-down strategies seem to be in many cases just "provisional", i.e. scientist only use them until bottom-up approaches become feasible. However, it is not completely certain that this happens/should happen in symbiosis research, for many times both strategies are combined (**Figure 3**). To understand the exact significance of bottom-up approaches in symbiosis research becomes a fundamental task for contemporary research in philosophy of science.

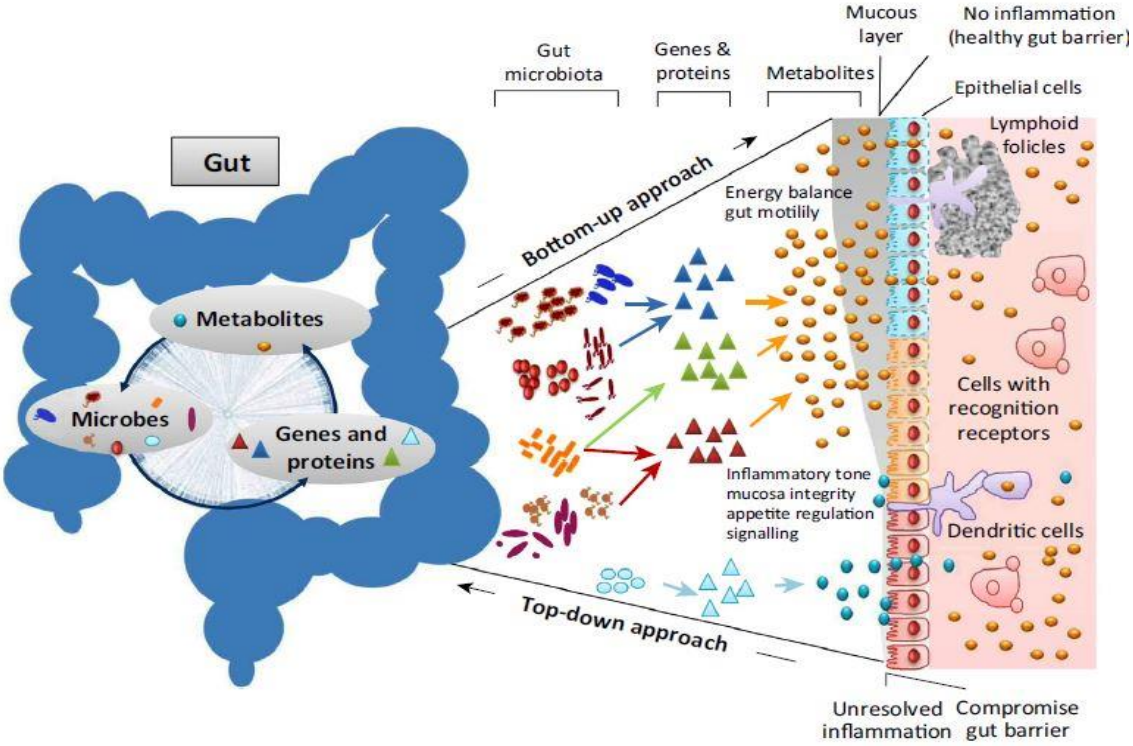

Figure 3. Combination of top-down and bottom-up approaches in contemporary microbiome research. In top-down approaches, it is studied how global properties of the system (e.g. degree of inflammation of the gut barrier) determine/affect its micro-components (e.g. genetic and microbial composition). In bottom-up approaches, on the contrary, it is studied how micro-components (e.g. microbial composition) affect global properties (e.g. the thickness of the mucus layer). These approaches are frequently combined in multispecies research, since both global properties affect micro-components, and vice-versa. The image has been taken from https://phys.org/news/2016-03-gut-microbiome-remarkably-stable.html.

Given the previous context, the purpose of this doctoral thesis is to study how new-mechanistic philosophy is being affected by symbiosis research, and how it may illuminate some aspects of contemporary science that have not being studied deeply enough yet. [4]

1.4. Justification for tackling this specific scientific and philosophical problem: What would we miss if we fail to investigate the philosophical principles that underlie contemporary symbiosis research?

The necessity of studying the philosophical assumptions –and their implications– that underlie contemporary research about multispecies symbiotic systems transcends the realm of philosophy of science. It derives from the deep implications that an underestimation –or an overestimation– of the philosophical grounding of these techniques may have for other applied disciplines such as contemporary evolutionary biology, or contemporary medicine. Disregarding or inadequately appreciating these implications may have consequences for:

- *Contemporary medicine – Human health*. The microbiome (a genuine symbiotic multispecies system) plays a fundamental role in human health and has been systematically related to diseases such as asthma, chronic diabetes, hypertension and inflammatory bowel disease, among others – probably including degenerative brain disorders, such as Parkinson (Ding et al. 2019). Contemporary medicine has designed treatments based on the principles of microbiome arrangement (including the fact that it is modelled by using network models), like the famous faecal transplantation, used to cope with chronic diarrhoea caused by *Clostridium difficile* infections. Despite the effectiveness of these treatments, we still lack a clear understanding of the principles that regulate microbiome arrangement, and most of these treatments are based on an inferred "inversed causality": if this treatment works, then it must be because it rearranges the microbiome (Hooks & O'Malley 2017). This type of inferences about the microbiome are potentially very promising, as they may lead to the development of new and necessary therapies against infectious diseases, particularly important in the days of antibiotic-resistance. However, they are also potentially dangerous, for the principles that regulate their functioning are still unknown, and they may cause a "health disaster" in the long term.

Given that my doctoral project investigates the deep epistemological and ontological assumptions and implications that underlie microbiome

[4] It is important to note, although in passing, that there are other lines of criticism to the new-mechanistic tradition that concentrate more on the way of defining the new-mechanistic commitments. For instance, some people have argued that new-mechanists have failed to provide a coherent notion of "mechanism", and that the term is used to define a heterogeneous class of phenomena across the sciences (Nicholson 2012; see also Woodward 2013). Yet other have contested the necessity of new-mechanistic philosophy on the basis that the tradition that new-mechanism criticises (Hempel's model of scientific explanation) can indeed be updated and made coherent with contemporary developments in philosophy of science (Díez 2014; Alleva et al. 2017). Even though these are solid criticisms against new-mechanism in philosophy of science, I will not consider them here, for the aim of the thesis is rather to investigate a set of biological phenomena that seem not to fit well with the type of phenomena that inspired new-mechanism.

research, its implications may be potentially useful for medicine, solving some of the contemporary questions about the universality of microbiome-based treatments, their scope, and their limits.

- *Contemporary evolutionary biology – Biodiversity preservation and agriculture*. Under the hologenome framework of evolution, the microbiome has been speculated to play a fundamental role in the evolution of animals and plants. Concretely, recent ground-breaking research in evolutionary biology has suggested that the adaptation to some specific dietary sources –e.g. strict herbivory in ruminants (see Gilbert 2019), or strict hematophagy in bats (see Mendoza et al. 2018)–, or to specific niches (e.g. Caribbean waters) may have been the result of a process of coevolution between the animal/plant genome, and the genetic composition of its microbiome. Multispecies symbiosis research has studied this type of co-evolutionary dynamics between hosts and their microbiome, with the aim of unveiling some of the specific details of animal-microbiome evolution. If this type of results can be generalised to other species, and to every adaption to rapid changing environmental and/or dietary conditions, it may potentially have a substantial use in programmes of biodiversity preservation (Royal Society 2018; Sergaki et al. 2018; Toju et al. 2018). Could the microbiome be selectively modified so that some endangered species can cope with the challenges of a rapidly changing climate? Could biodiversity be preserved by investing in programmes that, in addition to directly combating climate change, allow for more plastic and resistant animal and plant species on the basis of a modification of their microbiome? And, given the potential effects of the microbiome on agriculture, could some of these techniques be applied in that field? Is it possible to design resistant crops by selectively modifying their microbiome, rather than their genetic composition?

One of the most serious limitations of this type of evolutionary research is though that the scope of these results is limited, for we still have a proper and complete understanding of the kind of causal implications that can be derived from them. And, importantly, we currently ignore whether they can be studied under the principles of the new-mechanistic philosophy. My doctoral project aims at covering this gap by precisely providing a solid understanding of the epistemological assumptions that underlie this type of research –particularly, how they fit with the new-mechanistic philosophy–, thus uncovering its potential limits.

- *Contemporary philosophy – Epistemology*. As explained above, taking contemporary multispecies research seriously may radically change our view of some epistemological and ontological categories. Concretely, it may help understand better the nature of scientific modelling, insofar as my doctoral project will incorporate a whole new body of models that have been mostly disregarded in contemporary philosophy (Knuttila 2011; Frigg 2012; Green 2013; Serban & Green, forthcoming). This may prompt a new, wider view of the nature of biological knowledge, widening or radically changing

the scope of classical debates such as the debate about the nature of scientific explanation, the nature of causation, and the notion of emergence, whose study constitutes the main body of the project. Since all the new modelling techniques that are applied in the realm of contemporary biology are arguably explanatory, can lead to the discovery of a certain type of causal connections within multispecies systems, and are arguably the result of emergent properties (Osmanovic et al. 2018; Suárez & Triviño 2019, 2020; Green, forthcoming), studying each of these new techniques is highly promising for contemporary philosophy of science.

My doctoral project advances the field by posing new philosophical questions and provide answers to questions that are mostly neglected in contemporary philosophy of science. Are there forms of explanation that transcend the new-mechanistic paradigm? How do they work? Is it possible to see forms of causation that act top-down, rather than bottom up? How does symbiosis research illuminate these alternative forms of causation, and what does this tells us about causation?

Because of these issues, this doctoral project will have clear implications not only for contemporary philosophy of science, but also for contemporary science, in a way that can strongly advance our knowledge of the biological world.

2. Research goals of the thesis and guiding hypotheses

This doctoral thesis has four key research goals (RGs):

* **RG1.** To analyse the type of methodologies that underlie contemporary symbiosis research, with special emphasis on the type of mathematical tools that are used to articulate scientific explanations.

* **RG2**. To track the historical origins of the ideas that ground contemporary symbiosis research and relate them systematically to the type of philosophical assumptions that underlie contemporary research.

* **RG3.** To study the methodological differences between different types of symbiosis research, with focus on the differences between two-species models (e.g. host-parasite interactions), and multispecies models (e.g. the microbiome).

* **RG4.** To evaluate the epistemological consequences of the use of these tools in the context of the philosophical debate about scientific explanation, with special emphasis on how contemporary symbiosis research may limit some of the assumptions of new-mechanistic philosophy.

The project is grounded on two key guiding hypotheses (GHs), whose investigation constituted the core of the doctoral project:

*** GH1.** The pervasive use of holistic methodologies in contemporary symbiosis research challenges some of the most important philosophical assumptions that ground the new-mechanistic tradition in philosophy of science.

*** GH2**. The use of these new methodologies is rooted in a set of distinctive philosophical ideas about the concept of 'biological individuality' that contrast sharply with the ideas that support the application of mechanistic methodologies.

3. Research methodology applied in the thesis

During the doctoral project, the GHs have been tested in connection with historical and contemporary symbiosis research. The project has been grounded on two different research methodologies that have guided all the research activities.

(1) *Analysis of scientific practise*. The method of analysis of the scientific practice is quite common in contemporary philosophy of science (Nicholson & Dupré 2018). It basically consists in the careful examination of the ways in which scientists conduct their research, by placing special emphasis in the knowledge of the methodologies they apply and the reason(s) why they have chosen them. It has mostly been carried out by reading papers, although it is usually combined with direct talk to the scientists (e.g. by email exchange to ask them specific questions, meetings in conferences, personal discussion, etc).

(2) *Conceptual analysis*. Conceptual analysis is one of the oldest methods in the tradition of analytic philosophy (Soames 2003; Margolis & Laurence 2014). I have used it to study the implicit meaning of the concepts that biologists use in their research, how they are systematically related to each other, and which are their philosophically hidden implications. This method has enabled the doctoral candidate to better grasp the general implications (for epistemology and ontology) of the concepts that biologists use.

Each of these methodologies has been carried out in separate chronological moments, although both methodologies have been combined to a certain extent in the moment of writing the papers.

RESEARCH PERIOD 1. OCTOBER 2017 TO APRIL 2018

(Including a 3-months research stay at Egenis: The Centre for the Study of Life Sciences, University of Exeter, from January to April 2018)

During this period my research mostly consisted in the *analysis of the scientific practise*, with special emphasis on the study of the ontological implications of the methodological commitments that can be found in contemporary symbiosis research. I investigated the scientific literature, with emphasis on the research undertaken under the framework of the 'hologenome concept of evolution'. I read both papers that supported the hologenome hypothesis theoretically (Zilber-Rosenberg & Rosenberg 2008; Gilbert et al. 2012; Theis

et al. 2016) and empirically (Sharon et al. 2010; Brucker & Bordenstein 2013), with special interest in understanding the type of methodologies that were used in this type of research. At the same time, I also read papers that were critical with the hologenome hypothesis, both at the theoretical (Moran & Sloan 2015; Douglas & Werren 2016) and experimental levels (Chandler & Turelli 2014; Hester et al. 2016). Finally, I also read some classical works in history of biology that introduced the key concepts that my thesis was going to investigate (de Bary 1879; Pound 1893; Schneider 1897; Merezhkowsky 1905).

This research period was oriented towards understanding the type of ontological commitments at the core of multispecies research. Particularly, it aimed to disentangle the ontological basis of the scientific disagreement. In that vein, the type of methodology adopted was close to the type of methodology that Peter Godfrey-Smith has suggested for what he calls *philosophy of nature* (Godfrey-Smith 2009), and what Thomas Pradeu has characterised as philosophy *in* science.

At the end of this research period, the first paper of the doctoral dissertation had been published in *Symbiosis*, appearing in the number of October 2018.

RESEARCH PERIOD 2. MAY 2018 TO DECEMBER 2019

(Including a 3-months research stay at the Konrad Lorenz Institute for Evolution and Cognition Research, from March to June 2019)

In the last period of my research I applied conceptual analysis to the type of methodological commitments that I discovered in contemporary symbiosis research. Concretely, I studied two specific models that have been applied in contemporary symbiosis research, as I took them to be the most important for the overall goals of my doctoral research. The analysis was not aimed at understanding the models and their significance in the context of contemporary biology –this activity corresponded in any case to research period 2–, but at conceptually understanding their implications for contemporary philosophy. The goal was to understand to what extent the two models I had chosen (Blaser & Kirchner's equilibrium model, and Coyte et al.'s model of the microbiome) questioned some of the philosophical premises that grounded contemporary research in the new-mechanistic tradition.

The work in the research period was framed in the context of philosophy of science, as this discipline has been developed since it was first defended as an independent philosophical discipline due to the work of the Vienna Circle (Godfrey-Smith 2003; Díez & Moulines 2008; Rosenberg 2011).

At the end of this research period, I had published the two other papers that compose the doctoral dissertation, one appeared in *Synthese* in May 2019, and the other appeared in *Teorema: Revista Internacional de Filosofía*, in December 2019.

4. Research outcomes of the doctoral thesis: Description of the three papers

The doctoral research has given rise to three papers, each of which studies one of the dimensions that the doctoral project aimed to cover. The content of each of the papers,

how it relates to the RGs of the doctoral project, and how the three papers conform to each of the GHs is explained below.

4.1. The importance of symbiosis in philosophy of biology

The first paper, 'The importance of symbiosis in philosophy of biology: An analysis of the debate on biological individuality and its historical roots', appeared in the journal *Symbiosis*, volume 76, issue 2, pages 77-96, in October 2018 (first online: 28th April 2018). https://doi.org/10.1007/s13199-018-0556-1. It mainly contributes towards ***RG2**, as it studies the historical origins of some of the ideas that underlie contemporary symbiosis research, putting emphasis in the notion that contemporary disputes about the *legitimacy* of some types of symbiosis research –investigation of the 'hologenome hypothesis'– are in the end "old wine in new barrels".

The first part of the paper discusses the work of four biologists in late 19th and early 20th century (A. de Bary, R. Pound, A. Schneider, C. Merezhkowsky), to argue that their disagreements about the type of methodological practices they were using in their research were grounded on different views of 'biological individuality': while researchers such as de Bary, Schneider and Merezhkowsky had a "permissive" view, according to which some symbiotic assemblages could be considered biological individuals *and studied as such*, others (R. Pound) maintained a more limited view of the concept that discarded certain type of research strategies. Importantly, the paper strongly suggests that their disagreement was mainly *ontological*, i.e. it was based on different conceptions of what makes a "bunch of living stuff" a biological individual, versus what makes it a community of independent but interacting individuals.

The second part of the paper systematically reviews contemporary symbiosis research, when this is guided under the so-called 'hologenome framework of evolution'. It reviews the contemporary origins of this type of research (in the work of Lynn Margulis), as well as early criticisms (the work of John Maynard-Smith), to later show how the hypothesis has developed since it was explicitly formulated by biologists Ilana Zilber-Rosenberg and Eugene Rosenberg. This part of the review concentrates not only on how hologenome research has developed, but also on the type of criticism that it has given rise to. The paper systematically relates the disagreement about the usefulness/validity of the hologenome concept to disagreements about the nature of biological individuality, to quote:

> *current biological disputes are partially grounded in different philosophical assumptions, but concretely grounded in different conceptions about biological individuality* (2018: 93)

It goes further to argue that these disagreements are not different from the type of disagreements found in early symbiosis research, in late 19th/early 20th century, insofar as they are rooted in the same type of diverging philosophical assumptions, which are in the end ontological disagreements about "how the world should be sliced into pieces".

In this vein, the paper serves to pave the way for the rest of the doctoral project, which will concern investigating the type of philosophical assumptions that underlie contemporary symbiosis research.

The paper appeared in a *biology journal*, rather than in a philosophy journal, because while its main contribution was philosophical –to the literature in philosophy of biology (as the title and keywords suggest)– its intended target audience was mainly biological. On the one hand, the fact that the paper appeared in a biology journal proves the competence of the doctoral candidate to carry out the research he aims to carry out, as well as to discuss the type of philosophical implications that the thesis aims to investigate, for it proves his biological expertise. On the other hand, it also contributes to an ongoing scientific discussion, which despite not being the main goal of my doctoral research, is also an important goal in it.

4.2. Equilibrium explanation as structural non-mechanistic explanation

The second paper, 'Equilibrium explanation as structural non-mechanistic explanation: The case long-term bacterial persistence in human hosts', appeared in the journal *Teorema: Revista Internacional de Filosofía*, volume 38, issue 3, pages 95-120, in December 2019. http://philsci-archive.pitt.edu/id/eprint/16434. It mainly contributes towards ***RG3** and, partially, towards ***RG1**, as it investigates the study of two species symbiotic associations. The paper argues that *even* some studies of two-species symbiotic associations are non-mechanistic, thus contradicting the belief that the methodologies applied to study two-species systems need to be mechanistic.

The paper concentrates on the study of Blaser & Kirschner's equilibrium model. Their model was developed to explain why (and *how*) some symbiotic associations could persist long-term. Conventional population genetic knowledge suggests that symbiotic associations are inherently instable, for they would give rise to the evolution of cheaters that will disrupt the association. As this is so, the evolution of evolutionarily resilient symbiotic associations becomes almost impossible, and their existence would require explanation. Blaser & Kirschner's model covers that gap and provides a general framework to conceive long-term host-bacterial symbiotic associations. Importantly, it does so by relying on a mathematical development of the concept of evolutionary stable strategy *specifically applied to cases of symbiosis* –the concept of evolutionary stable strategy is conventionally applied to individuals of the same population. And, second, Blaser & Kirschner do not rely on *any* mechanism to develop their model, which only concentrates on potential structural features of any host-symbiont association. Their key contribution is thus to explore the space of possible biological states that may lead to the appearance of a long-term host-bacterial association.

Grounded on this research, our paper argues that Blaser & Kirschner's model is non-causal, if causality is conceived in mechanistic terms. Our key argument is their model relies on the concept of evolutionary stable strategy, and thus it only explores the possible equilibrium states that would make host-bacterial symbiotic associations persistent in the long-term, and in three different levels (molecularly, ontogenetically, evolutionarily). Instead of relying on the possible mechanistic structure of host-bacterial symbiotic associations, Blaser & Kirschner's model relies exclusively on their mathematical properties, and thus the model has enough generality for being applies to every possible symbiotic system (rather than to a concrete pair of species). The key lesson that our paper provides in that even pair host-bacterial symbiotic associations are

sometimes studied mathematically, which strongly questions the applicability of new-mechanism in contemporary symbiosis research.

4.3. Explaining the behaviour of random ecological networks

The third paper, 'Equilibrium explanation as structural non-mechanistic explanation: The case long-term bacterial persistence in human hosts', appeared in the journal *Synthese*, in May 2019 (no volume or issue assigned yet). https://doi.org/10.1007/s11229-019-02187-9. It mainly contributes towards ***RG4** and, partially, towards ***RG1**. The paper investigates the type of modelling techniques that are used for the study of the behaviour of multispecies symbiotic systems. Concretely, it focuses on the microbiome (a multispecies symbiotic system) and studies the type of modelling techniques that contemporary biologists are using to study some of its more salient ecological properties (stability behaviour).

The paper studies the philosophical assumptions that underlie Coyte et al.'s model of the microbiome. Their model relies on network analysis to study the ecology of the microbiome, concretely, its stability behaviour. Coyte et al.'s model serves to prove that what makes the microbiome stable is the high degree of competition among the different bacterial species that compose it, rather than the high degree of interspecies cooperation (as other ecological models had previously suggested). To prove this, Coyte et al. build a mathematical (network) model of the microbiome and study how it will respond to different perturbations given certain initial conditions. Their model, importantly, is purely mathematical, and takes the microbiome holistically, rather than de-composing it in its component parts.

Grounded on that, in our paper, we make a philosophical analysis of Coyte et al.'s methodology to better understand its implications for contemporary philosophy of science. Our paper holds an important promise for future research, for it proves three theses: (1) in the line of the second paper, and in agreement with GH1, it contradicts the universality of the new-mechanistic philosophy by denying that every biological explanation must be causal, where causality is understood in the very strict mechanistic terms (bottom-up, studied by de-composing the system in its component parts); (2) it suggests a view of scientific explanation as a *pluralist*, *integrative* activity, where the model of the mechanism is combined with a mathematical structure that limits the range of possible states of the system. This point, far from being trivial, holds a promise for developing future, original research on the type of elements that make this type of explanations genuinely explanatory.

4.4. How the papers confirm/reject the guiding hypothesis

The three papers that constitute the body of the doctoral dissertation confirm well GH1, namely:

> *** GH1.** The pervasive use of holistic methodologies in contemporary symbiosis research challenges some of the most important philosophical assumptions that ground the new-mechanistic tradition in philosophy of science.

My research has particularly proven that the demand that biological systems must be decomposed to provide genuine explanations has been completely challenged. Contrary to that, contemporary symbiosis research proves that the features of some biological system are explained only holistically and, more importantly, *they can only be explained holistically*. In addition to that, my research has proven that the properties of some biological system need to be studied top-down, rather than bottom-up (as new-mechanists demand), which opens an important question about causality (see next section).

My research, though, has been made coherent with GH2, although it has not completely proven it. To repeat, GH2 asserted:

> *** GH2**. The use of these new methodologies is rooted in a set of distinctive philosophical ideas about the concept of 'biological individuality' that contrast sharply with the ideas that support the application of mechanistic methodologies.

It cannot be denied that my doctoral dissertation has proven that part of the divergence between contemporary symbiosis research and other forms of research in contemporary biology rest on different ideas about how to set the boundaries of biological individuals (see especially paper 1). However, there is still some room left to confirm the truth of GH2. Particularly, the specific ontology that justifies the application of different methodologies requires further investigation. A recent draft, jointly written with Álvaro Moreno, provides stronger support to GH2. However, since it has not been accepted yet, it is not included in the dissertation, and thus the dissertation provides support for GH2 *without completely confirming its truth*.

5. Open questions that arise from the doctoral research

Although the doctoral project provided answers to the questions it aimed to tackle, it also left several open questions whose investigation is left for future research.

Firstly, the dissertation served to prove that some contemporary biological research (symbiosis research) is at odds with the new-mechanistic paradigm in philosophy of science. It showed that this paradigm could be questioned not only in multispecies research, but also, and more importantly, in single host-parasite research. Now, this clearly opens an important question: if biologists are currently designing research strategies that are at odds with the new-mechanistic paradigm and that, I argued, are perfectly explanatory, what makes these activities explanatory activities? New-mechanism provides a sold framework to justify why certain types of scientific activities are explanatory. This thesis only proves that new-mechanistic philosophers have systematically ignored a part of scientific research that relies on different principles to build its explanations. However, *it has not developed a systematic account of why these contemporary methods of biological research are also explanatory* (although see Díez 2014, for a proposal). Providing an answer to that question is an open task for future research, one that requires fundamental attention.

Secondly, in section 1.2. I explained how the new-mechanist philosophy provides a useful account of how some scientific methodologies are able to track causal relationships in the world. Importantly, these methodologies are often used as a basis to

define a philosophically enriching conception of "causality". The research that I present in this doctoral dissertation severely questions that every scientific methodology traces/investigates the type of causal relationships *that new-mechanist philosophers define*. Despite this, it is undeniable that contemporary symbiosis research has discovered systematic relationships *of some sort*. A key question that has been answered in my doctoral research concerns the later point: what type of systematic relationship are being studied in contemporary symbiosis research? This question is crucial, for it can guide a completely new research into the concept of *causality*. Is the category of "causality" wider than new-mechanistic philosophers believe? Is it possible to argue that there are different degrees of causality? Can causality be a systemic property that, as such, is responsible of the behaviour of the parts that compose the system, rather than the other way around? If this is so, can we restate the concept of "downward causation"? On which grounds? These questions require an answer, and my doctoral project has only opened them.

Finally, the project results raise a very important ontological question. I said that the doctoral project had proven that some biological systems can only be known holistically, for their systemic properties disappear as soon as the system is de-composed to study the behaviour of each of its parts. In other words, how two parts interact with each other is irrelevant to know how the multispecies symbiotic system will act, as its behaviour becomes unpredictable. This fact, of course, raises a substantial question about *emergence*. Do symbiotic systems manifest emergent properties? Does their existence suppose an emergent jump in nature? How can their properties be characterised? Why is the behaviour of these systems non-predictable?

REFERENCES

Alleva, K., Díez, J. A., & Federico, L. (2017) 'Models, theory structure and mechanisms in biochemistry: The case of allosterism.' *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 63: 1-14.

Baedke, J (2019) 'O organism, where art thou? Old and new challenges for organism-centered biology.' *Journal of the History of Biology* 52(2): 293-324.

Bapteste, E., Huneman, P. (2018) **'**Towards a Dynamic Interaction Network of Life to unify and expand the evolutionary theory.' *BMC Biol* 16: 56.

Batlle, E., & Wilkinson, D. G. (2012) 'Molecular mechanisms of cell segregation and boundary formation in development and tumorigenesis.' *Cold Spring Harbor perspectives in biology* 4(1): a008227.

Bechtel, W., A. Abrahamsen (2005) 'Explanation: A mechanist alternative.' *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 421–441.

Bechtel, W., Richardson, R. C. (1993) *Discovering complexity: Decomposition and localization as scientific research strategies*. Cambridge, MA: The MIT Press.

Blair, J., Webber, M., Baylay, A. et al. (2015) 'Molecular mechanisms of antibiotic resistance.' *Nat Rev Microbiol* 13: 42–51.

Blaser, M. J., D. Kirschner (2007) 'The equilibria that allow bacterial persistence in human hosts.' *Nature* 449: 843–849.

Bourrat, P. (2018) 'Have Causal Claims About the Gut Microbiome Been Over-Hyped?' *BioEssays* 40:12.

Brucker, R.M., S.R. Bordenstein (2013) 'The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus Nasonia.' Science: 1240659.

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

Chiu, L., S.F. Gilbert (forthcoming) 'Niche construction and the transition to herbivory: Phenotype switching and the organization of new nutritional modes'. In: H. Levine, M. Jolly, P. Kulkarni, V. Nanjundiah (eds.) *Phenotypic switching: Implications in Biology and Medicine*.

Clarke, E. (2010) 'The problem of biological individuality' Biol Theory 5(4): 312-325.

Coyte, K. Z., Schluter, J., & Foster, K. R. (2015). 'The ecology of the microbiome: Networks, competition, and stability.' *Science* 350(6261): 663–666.

Craver, C. F. (2007) *Explaining the brain*. New York: Clarendon Press.

Craver, C. F. (2014) The ontic account of scientific explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, and A. Hüttemann (eds.). *Explanation in the special sciences: The case of biology and history*. Springer Verlag. pp. 27–52.

Craver, C. F., L. Darden (2013) *In search for mechanisms: Discovery across the life sciences*. Chicago: University of Chicago Press.

Crews, L., Masliah, E. (2010) 'Molecular mechanisms of neurodegeneration in Alzheimer's disease'. *Human Molecular Genetics* 19 (R1): R12–R20.

De Bary A (1879) Die Erscheinung der Symbiose. Verlag von Karl J, Trübner.

Deulofeu, R., J. Suárez (2018) When mechanisms are not enough: The origin of eukaryotes and sceintific explanation. In A. Christian, D. Hommen, N. Retzlaff, G. Schurz (eds.) *Philosophy of Science*. European Studies in Philosophy of Science, vol. 9. Springer, Cham.

Díez, J.A. (2014) Scientific w-explanation as ampliative, specialized embedding: A neo-Hempelian account. *Erkenntnis* 79: 1413–1443.

Díez, J.A., C.U. Moulines (2008) *Fundamentos de Filosofía de la Ciencia*. Barcelona: Ariel.

DiFrisco, J. (2017) 'Kinds of biological individuals: Sortals, projectability and selection'. *Br J Philos Sci.* 70(3): 845-875.

Ding, R., W. Goh, R. Wu, et al. (2019) 'Revisit gut microbiota and its impact on human health and disease.' *Journal of Food and Drug Analysis* 27(3): 623-631.

Douglas, A. E., J.H. Werren (2016) 'Holes in the hologenome: why host-microbe symbiosis are not holobionts'. mBio 7(2): e02099-15.

Dupré, J., M. A. O'Malley (2009) 'Varieties of living things: Life at the intersection of lineage and metabolism' PTPBio 1(3).

Frigg, R. (2012) 'Models in science'. *The Stanford Encyclopaedia of Philosophy*. https://plato.stanford.edu/entries/models-science/. Accessed 5 December 2019.

Giere, R. N. (1999) *Science Without Laws*. Chicago: University of Chicago Press.

Gilbert, S. (2019) 'Developmental symbiosis facilitates the multiple origins of herbivory.' Evolution & Development. https://doi.org/10.1111/ede.12291. Accessed 29 November 2019.

Gilbert, S., J. Sapp, A. Tauber (2012) 'A symbiotic view of life: We have never been individuals', Quarterly Review in Biology 87(4): 325-341.

Gilbert S., E. Rosenberg, et al. (2017) The holobiont with its hologenome is a level of selection in evolution. In SB Gissis, E Lamm & A Shavit (eds.) *Landscapes of collectivity in the life sciences*. London: The MIT Press: 305-324.

Gissis, S.B., E. Lamm, A. Shavit (2018) *Landscapes of collectivity in the life sciences*. Cambridge MA: The MIT Press.

Glennan, S. (1996) 'Mechanisms and the nature of causation.' *Erkenntnis*, 44(1): 49–71.

Glennan, S. (2002) 'Rethinking mechanistic explanation.' *Philosophy of Science* 69 (S3): S342–S353.

Glennan, S., & Illari, P. (2017) *The Routledge handbook of mechanisms and mechanical philosophy*. Taylor & Francis.

Godfrey-Smith, P. (2003) *Theory and Reality: An Introduction to the Philosophy of Science*. University of Chicago Press.

Godfrey-Smith, P. (2009) Darwinian populations and natural selection. Oxford: Oxford University Press.

Green, S. (2013) 'When one model is not enough: Combining epistemic tools in systems biology' Stud Hist Philos Sci 44(2): 170-180.

Green, S. (forthcoming). 'Cancer beyond genetics: On the practical implications of downward causation.' In D. S. Brooks, J. DiFrisco, & W. C. Wimsatt (Eds). Biological Levels: Composition, Scale and Evolution in Complex Systems. MIT Press.

Green, S., N. Jones (2016) 'Constraint-based reasoning for search and explanation: Strategies for understanding variation and patterns in biology.' *Dialectica* 70(3): 343-374.

Hooks, K. B., O'Malley, M. A. (2017) 'Dysbiosis and its discontents.' *mBio* 8(5): e01492-17.

Huneman, P. (2010) 'Topological explanations and robustness in biological sciences.' Synthese 177: 213–245.

Huneman, P. (2018) 'Outlines of a theory of structural explanation.' Philosophical Studies 175 (3): 665–702.

Issad, T., Malaterre, C. (2015) 'Are dynamic mechanistic explanations still mechanistic?' In P. A. Braillard & C. Malaterre (eds.) *Explanation in biology: An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 265–292). Dordretcht: Springer.

Kaplan, D. M., & Craver, C. F. (2011) 'The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective.' *Philosophy of Science* 78(4): 601–627.

Knuuttila, T. (2011) 'Modelling and representing: An artefactual approach to model-based representation' *Stud Hist Philos Sci* 42: 262-271.

Krickel, B. (2018) *The Mechanical World: The Metaphysical Commitments of the New Mechanistic Approach*. Dordrecht: Springer.

Laplane, L., P. Mantovani, et al. (2019) 'Why Science Needs Philosophy?' PNAS 116(10): 3948-3952.

Lidgard, S., L. K. Nyhart (2018) Biological Individuality: Integrating Scientific, Philosophical, and Historical Perspectives. University of Chicago Press.

Lloyd E (2017) Holobionts as units of selection: Holobionts as interactors, reproducers, and manifestors of adaptation. In: Gissis SB, Lamm E, Shavit A (eds) *Landscapes of collectivity in the life sciences*. The MIT Press, London, pp 351–367

Lloyd, E., M. J. Wade (2019) Criteria for holobionts from community genetics. Biol Theor. https://doi.org/10.1007/s13752-019-00322-w.

Lynch, K. E., E. C. Parke, M. A. O'Malley (2019) How causal are microbiomes? A comparison with the Helicobacter pylori explanation of ulcers. Biol Philos 34: 62.

Machamer P, L Darden, C. Craver (2000) Thinking about mechanisms. Philosophy of science, 67(1): 1-25.

Margolis E., S. Laurence (2014) Concepts. In EN Zalta (ed.) The Stanford Encyclopaedia of Philosophy. https://plato.stanford.edu/entries/concepts/. Accessed 6 December 2019.

Mendoza, M. L. Z., Z. Xiong, et al. (2018) 'Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat.' *Nature Ecol Evol* 2: 659–668.

Merezhkowsky C (1905) 'Über Natur und Ursprung der Chromatophoren imPflanzanreiche.' *Biologisches Centralblatt* 25:593–604.

Militello, G., A. Moreno (2018) 'Structural and organizational conditions for being a machine.' *BioPhilos*, 33, 35.

Mitchell, S. D. (2003) *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

Moran, N., D. B. Sloan (2015) The Hologenome Concept: Helpful or Hollow? PLoSBiol 13 (12): e1002311.

Nicholson, D. J. (2012) 'The concept of mechanism in biology.' *Stud Hist Phil Biol Biomed Sci* 43 (1): 152–163.

Nicholson, D. J. (2013) 'Organism ≠ Machines.' *Stud Hist Phil Biol Biomed Sci*, 44, 669-678.

Nicholson, D. J. (2018) 'Reconceptualizing the Organism: From Complex machine to Flowing Stream'. In: D. J. Nicholson and J. Dupré (eds). *Everything Flows: Towards a Processual Philosophy of Biology* (pp. 139-166). Oxford: Oxford University Press.

Nicholson, D.J., J. Dupré (2018) E*verything Flows: Towards a Processual Philosophy of Biology*. Oxford: Oxford University Press.

O'Malley, M., D.Skillings (2018) 'Methodological strategies in microbiome research and their explanatory implications.' *Persp on Sci* 26(2): 239-265.

Osmanovic, D., D.A. Kessler, et al. (2018) 'Darwinian selection of host and bacteria supports emergence of Lamarckian-like adaptation of the system as a whole.' *Biol direct* 13(1): 24.

Pepper, J. W., M. D. Herron (2008) 'Does biology need an organism concept?' *Biol Rev Camb Philos Soc* 83(4): 621-627.

Pound R (1893) 'Symbiosis and mutualism.' *Am Nat* 27(318): 509–520-

Pradeu, T. (2012) *The Limits of the Self: Immunology and Biological Identity*. Oxford: Oxford University Press.

Pradeu, T. (2019) 'Philosophy of Biology: Immunology and individuality.' *eLife*. https://elifesciences.org/articles/47384. Accessed 1 December 2019.

Rosenberg, A. (2011) *Philosophy of Science: A Contemporary Introduction*. London: Routledge.

Rosenberg, E., I. Zilber-Rosenberg (2014) *The Hologenome Concept: Human, Animal, and Plant Microbiota*. Dordrecht: Springer.

Roughgarden, J., S. Gilbert, et al. (2018) Holobionts as units of selection and a model of their population dynamics and evolution. Biological Theory 13(1): 44–65.

Salmon, W. (1984) *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Schneider A (1897) 'The phenomena of Symbiosis.' *Minnesota Botanical Studies* 1(9): 923–948.

Serban, M., S. Green (forthcoming) 'Biological robustness: Design, organization and mechanisms.' In Holm, S.H. & Serban, M. (Eds). *Living Machines? Philosophical Perspectives on the Engineering Approach in Biology*. Routledge.

Sergaki, C., B. Lagunas, et al. (2018) 'Challenges and approaches in microbiome research: From fundamental to applied.' Frontiers in Plant Science 9: 1205.

Sharon, G., D. Segal et al. (2010). 'Commensal bacteria play a role in mating preference of Drosophila melanogaster.' PNAS USA 107(46): 20051-20056.

Stencel, A., D. Proszewska (2017) 'How Research on Microbiomes is Changing Biology: A Discussion on the Concept of the Organism'. Foundations of Science 23(4): 603-620.

Stencel, A., D. M. Wloch-Salamon (2018) 'Some theoretical insights into the hologenome theory of evolution and the role of microbes in speciation.' *Theory Biosci*. 137(2): 197-206.

Suárez, J., V. Triviño (2019) 'A metaphysical approach to holobiont individuality: Holobionts as emergent individuals.' *Quaderns de Filosofia* 6(1): 59-76.

Suárez, J., V. Triviño (2020) 'What is a hologenomic adaptation? Emergent individuality and inter-identity in multispecies systems.' *Frontiers in Psychology*. Forthcoming.

Theis, K.R., N.M. Dheilly, et al. (2016) 'Getting the hologenome concept right: An eco-evolutionary framework for hosts and their microbiomes'. mSystems 1(2): e00028-16.

The Royal Society (2018) Conference report: 'The microbiome: human medicine and agriculture in a microbial world'. https://royalsociety.org/~/media/events/2018/10/tof-microbiome/Transforming%20our%20future%20-%20The%20microbiome%20conference%20report.pdf?la=en-GB. Accessed 7th December 2019.

Toju, H., K.G. Peay, et al. (2018) 'Core microbiomes for sustainable agroecosystems.' *Nature plants* 4: 247-257.

Yuan, J., Lipinski, M., Degterev, A. (2003) 'Diversity in the Mechanisms of Neuronal Cell Death'. *Neuron* 40(2): 401-413.

Woodward, J. (2013) ‘Mechanistic explanation: Its scope and limits.’ *Aristotelian Society Supplementary* 87 (1): 39–65.

Zilber-Rosenberg, I., E. Rosenberg (2008) ‘Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution.’ FEMS Microbiology Reviews 32(5): 723-735.

II. ANNEXES

1. Paper 1: ‘The importance of symbiosis in philosophy of biology: An analysis of the debate on biological individuality and its historical roots’. Symbiosis 76/2: 77-96. doi: https://doi.org/10.1007/s13199-018-0556-1.

2. Paper 2: ‘Equilibrium explanation as structural non-mechanistic explanation: The case long-term bacterial persistence in human hosts’. Teorema 38/3: 95-120.

3. Paper 3: ‘Explaining the behaviour of random ecological networks: The stability of the microbiome as a case of integrative pluralism’. Synthese, online first: https://doi.org/10.1007/s11229-019-02187-9.

Symbiosis (2018) 76:77–96
https://doi.org/10.1007/s13199-018-0556-1

'The importance of symbiosis in philosophy of biology: an analysis of the current debate on biological individuality and its historical roots'

Javier Suárez[1]

Received: 9 December 2017 / Accepted: 13 April 2018 / Published online: 28 April 2018

Abstract

Symbiosis plays a fundamental role in contemporary biology, as well as in recent thinking in philosophy of biology. The discovery of the importance and universality of symbiotic associations has brought new light to old debates in the field, including issues about the concept of biological individuality. An important aspect of these debates has been the formulation of the hologenome concept of evolution, the notion that holobionts are units of natural selection in evolution. This review examines the philosophical assumptions that underlie recent proposal of the hologenome concept of evolution, and traces those debates back in time to their historical origins, to the moment when the connection between the topics of symbiosis and biological individuality first caught the attention of biologists. The review is divided in two parts. The first part explores the historical origins of the connection between the notion of symbiosis and the concept of biological individuality, and emphasizes the role of A. de Bary, R. Pound, A. Schneider and C. Merezhkowsky in framing the debate. The second part examines the hologenome concept of evolution and explores four parallelisms between contemporary debates and the debates presented in the first part of the essay, arguing that the different debates raised by the hologenome concept were already present in the literature. I suggest that the novelty of the hologenome concept of evolution lies in the wider appreciation of the importance of symbiosis for maintaining life on Earth as we know it. Finally, I conclude by suggesting the importance of exploring the connections among contemporary biology, philosophy of biology and history of biology in order to gain a better understanding of contemporary biology.

Keywords Symbiosis · History of biology · Philosophy of biology · Biological individuality · Hologenome · Holobiont · Units of selection

That symbiosis is a universal phenomenon in our planet is something that does not escape the attention of biologists. Organisms of different species constantly engage with each other in various types of associations, amongst which symbiosis–the persistent relationship among individuals of different species (Paracer and Ahmadjian 2000)– stands out as an essential phenomenon for the maintenance of life on Earth as we know it. For instance, it is widely acknowledged that the bodies of most animals contain an important number of bacterial partners, which sometimes even leads to the duplication of the number of their own cells (Huttenhower et al. 2012; Relman 2012; McFall-Ngai et al. 2013; McFall-Ngai 2015). Furthermore, the important role of symbionts for the physiology and normal development of their hosts is generally recognized and widely supported by current biological evidence (Gilbert and Epel 2009; Brucker and Bordenstein 2012, 2013; Sommer and Bäckhed 2013; Rosenberg and Zilber-Rosenberg 2014; McFall-Ngai 2015). Finally, the importance of symbiosis in some events of speciation has been recently explored and it is currently gaining empirical support (Jaenike et al. 2010; Brucker and Bordenstein 2012, 2013; Gontier 2015; Lipnicki 2015).

The acknowledgment of the importance of symbiosis for the maintenance of life on Earth, as well as the universality of the phenomenon, has recently led philosophers of biology to question the definition of some of the most important concepts in the field. Particularly important, the

✉ Javier Suárez
javier.suarez@ub.edu; jsuar3b@gmail.com

[1] Department of Philosophy, Universitat de Barcelona, C/ Montalegre 6-8, E–08001 Barcelona, Spain

monogenetic uniqueness of organisms as well as the boundaries between organisms and their environment have been questioned, challenging some of the traditional definitions of biological individuality. Should the symbiotic microorganisms that reside within the bodies of animals and plants be considered parts of a holistic unit that encompasses the host and its symbionts or, on the contrary, should those microorganisms be considered independently from the host? If they should be considered parts of the host, forming a higher level entity, what is the metaphysical status of this higher level entity? Is it itself a biological individual, or an ecological community of different independent individuals? Can symbiotic assemblages be considered units of selection, i.e. objects that form populations that evolve following Darwinian dynamics? If so, how does this affect the concept of heredity? And how do symbiotic associations evolve through time? As McFall-Ngai et al. have summarized: "[t]hese new data are demanding a re-examination of the very concepts of what constitutes a genome, a population, an environment, and an organism" (McFall-Ngai et al. 2013: 3234).

Those basic philosophical questions are not completely new, and their origin can be traced back in time to the first moments in which symbiosis thinking began to flourish, and the definition of the concept was not clear.[1] For instance, at the end of the nineteenth century a popular trend in biology started identifying symbiosis with "mutualism" (Martin and Schwab 2012). This attitude led those authors to emphasize the existence of a "shared dependency" among partners (physiological, morphological), in which the biological individuality of these partners might be sacrificed in benefit of the "bigger entity". This, for example, was the position of Albert B. Frank in his early research on mycorrhizas in 1885 (Frank 1885 [Frank 2005]; Trappe 2005). However, if on the contrary most symbionts are interpreted as parasites, this would led to emphasize the individuality of the symbionts that engage in the relationship, suggesting that they will engage in the relationship for their own benefit, not losing their autonomy. This position was vigorously defended by Roscoe Pound (1893). Their disagreement suggests that the conception of symbiosis that one holds has consequences for how to conceive biological individuality.

This review has two purposes: first, it aims to analyse the influence of symbiosis thinking in recent philosophy of biology, particularly reflecting how it has influenced the debates about the boundaries and constitution of the biological individual, as well as the debates about the units of selection; second, it aims to uncover the historical roots of the relation between the concept of symbiosis and the philosophical controversy about what constitutes a biological individual.[2]

The review will be divided into two parts. The first part, historically oriented, will introduce the concept of symbiosis and analyse its conceptual evolution since it was first proposed by Anton de Bary in 1879. The emphasis of this section will be put in how the concept of symbiosis did already question, since its original formulation, the boundaries and the constitution of biological individuals. The second part of the paper will be centred on recent developments in microbiology, paying special attention to the hologenome concept of evolution, and how those developments have affected current debates on the notion of what constitutes a biological individual, as well as its connection with the debate about the units of selection. I argue that most of the philosophical issues raised by the hologenome concept of evolution were already present in the original debates about symbiosis, and I try to uncover their historical roots, drawing four parallelisms between past research on symbiosis and the research done in the light of the hologenome concept. Finally, I suggest that the recent awareness of the philosophical significance of symbiosis originates from three facts: first, the appreciation on the universality of the phenomenon, which derives from the development of new techniques to identify the presence of microorganisms in the body of multicellular organisms (microbiomics); second, its importance for sustaining life as we know it, including the role that symbionts play in the *physiology* and *development* of multicellular organisms; third, the consideration of some symbiotic assemblages –holobionts– as units of selection, which caught the attention of philosophers who were previously not so interested in the phenomenon of symbiosis.

1 Part I. The historical roots of the concept of symbiosis – philosophical implications

This part of the review explores four main ideas relating to the concepts of symbiosis and biological individuality. The genesis of these ideas will be traced back to the authors that first proposed them. I will begin by considering the work of Anton de Bary, who first considers symbiosis as a separate biological phenomenon, naming it, and characterizing

[1] Many authors still argue that the definition of symbiosis are not clear at present, e.g. Wilkinson (2001), Douglas (2010: 4), Martin and Schwab (2013).

[2] The review is not about the problem of biological individuality and how different biologists and philosophers have conceived the topic; rather, this review is about the relation between symbiosis and certain dimensions of the problem of biological individuality –the boundaries and composition of the biological individual and the units of selection. However, the reader must at least take into account that three different notions of biological individual will be considered, especially in part 2: biological individuals as functionally integrated units, biological individuals as units of selection and biological individuals as bounded units (with clear physical boundaries, such as a membrane). Readers interested in the philosophical problem of biological individuality might refer to Wilson and Barker (2013), Bouchard and Huneman (2013), Pradeu (2016a), DiFrisco (2017) and Lidgard & Nyhart (2017: 17-63).

its specific properties. At this time, I will argue, symbiosis was already understood to challenge received ideas on the physiological boundaries of the individual; and yet, symbiosis was not then clearly distinguished from other phenomena, e.g. biological "sociality" (section 1). Second, I will consider Roscoe Pound's criticism of a symbiosis understood as mutualism, and I will introduce the arguments he presents to justify his opposition, paying special attention to his reliance on the concept of "struggle for life" (section 2). Third, I will analyse Albert Schneider's "The Phenomena of Symbiosis", as the first systematization of the concept and, more importantly for the purposes of this paper, the first moment in which symbiosis was understood as a phenomenon that might evolve over time, and which could be analysed independently of the organisms that interact symbiotically. I argue that Schneider supposes the first important step in considering symbiotic assemblages as units of selection (section 3). Finally, I argue that the last step for considering symbiotic assemblages as evolutionary individuals (i.e. as questioning the conventional frontiers of the evolutionary individual, of the entity that "struggles for life") was accomplished by Constantin Merezhkowsky,[3] when he hypothesized about the symbiotic origin or chloroplasts, thus creating the conceptual possibility of imagining a hereditary symbiosis (section 4).[4]

1.1 Anton de Bary (1831–1888)

The introduction of the term "symbiosis" in biology is usually credited to Anton de Bary, who originally used it for the first time in the history of biology in his speech to the Association of German Naturalists and Physicians, "*Die Erscheinung der Symbiose*", the "phenomenon of symbiosis" (de Bary 1878 [2016]; Oulhen et al. 2016).[5] Nonetheless, one year before de Bary's lecture, Albert B. Frank had introduced the term "*Symbiotismus*", to designate those "cases where two different species live on or in one another" (Frank 1877, quoted in Sapp 1994: 6). When they first used the term, both Frank and de Bary were interested in the study of lichens, whose dual nature had been hypothesized ten years before by Simon Schwendener (Honegger 2000; Egerton 2015; Gontier 2016a). For Schwendener, the dual nature of lichens was understood as a relationship where the fungus is in control of the algae, that it uses to obtain its nutrients, but also, and most importantly, as a new biological individual: "the organisms are so intrinsically and reciprocally connected that through their penetration and merging, they constitute new plants *with a clear individual character*" (1868, quoted in Gontier 2016a; emphasis added). De Bary, drawing upon those observations plus the experimental results that demonstrated that the dual nature of lichens was not merely a fiction of Schwendener –the two elements that constitute the lichen were separated for the first time in 1876, and by 1877 it was already possible to synthetize lichens in the lab by merging algae with fungal spores (Stahl 1877, referred in de Bary 1878 [2016]; Sapp 1994; Egerton 2015: 104, 106; Gontier 2016a: 276)–decided to refer to "the living together of differently named organisms" by the term "symbiosis" (1878 [Oulhen et al. 2016: 133]).

In his original lecture, de Bary emphasizes two aspects of symbiotic relationships: first, the different degrees of dependency that the partners in a symbiotic relationship sometimes generate with respect to each other; second, the different kind of effects that can be generated as a consequence of the symbiotic association. With respect to the latter point, by the time when de Bary coined the concept of "symbiosis", Pierre-Joseph van Beneden's classification of the different types of associations between organisms in mutualism, commensalism and parasitism, had become very popular (van Beneden 1876), and de Bary would precisely use that classification to better capture the nature of symbiotic phenomenon, a phenomenon of which, in his words: "[p]arasitism, mutualism and *lichenism* are special cases" (1878 [Oulhen et al. 2016: 136]; emphasis added).[6] With respect to the former, de Bary's speech is predominantly dedicated, in almost its totality, in explaining the types of dependencies among partners, including, especially, the morphological and physiological effects that symbiosis can cause the individuals that are interacting symbiotically. Interestingly, he decided to include *lichenism* as a distinct type of symbiotic relationship together. Why, then, is lichenism different to mutualism and commensalism?

Most of de Bary's paper is dedicated in explaining the association between *Azolla* and *Anabaena*, *Nostoc* and *Cycas*, and the fungi and algae that constitute lichens. In fact,

[3] There are alternative ways to spell his name (e.g. Merezhkovski, Mérejkovski, Mereschkovsky). I use the spelling that appears in Sapp et al. (2002).

[4] My historical focus is selective and not exhaustive, since I aim to compare four parallelisms between the historical development of the concept of symbiosis and the recent developments of the concept of holobiosis. The reconstructed history I will present will reflect this interest. For the readers who are interested in seeing different historical reconstructions see Sapp (1994), Paracer & Ahmadjian (2000: 231-238), Wilkinson (2001), Peacock (2011), Martin and Schwab (2012), Egerton (2015), Carrapiço (2015), Gontier (2015, 2016a), Zook (2015).

[5] Frank N. Egerton, however, in his review paper on the history of symbiosis studies dedicates the first section to studies of symbiotic phenomena that appeared before the concept of "symbiosis" was introduced (2015: 81–90). He goes as far as to Herodotus, Aristotle and Theophrastus. Despite the interest of their research, as far as this review is about the philosophical implications of the concept and its relation to other philosophical concepts, I have chosen to begin with de Bary's account.

[6] Parasitism was known while before van Beneden, but parasites (including those that we might call nowadays *microorganisms*, Pasteur's germs) were basically considered as pursuing their own interests, thus necessarily damaging the other in a context of struggle for life (e.g. Spencer 1899; cf. Sapp 1994: 25–28). Precisely, what is innovative about van Beneden's work was that he was the first in: (1) identifying the existence of an important number of associations among organisms that are not parasitic, a discovery that of course had historical precedents; (2) classifying the different types of biological associations in virtue of their effects in a systematic way, which is also conceptually different from previous views on the economy of nature (Egerton 2015: 84).

de Bary seemed to perceive something particular in those associations, which is the reason why he asserted:

> "When we observe more closely the phenomena described above, we find in the azollas and the cycads as well as in lichens, intimate associations of different species but never an organization that fits one of the categories described at the beginning of this study. For the reasons that I have already explained [see below], we cannot strictly speak of commensalism or parasitism" (1878 [Oulhen et al. 2016: 135])

Furthermore, he also discards the hypothesis that those associations might be considered simply as cases of mutualism: "[i]t is however doubtful that there are mutual advantages to the partners. We can definitely say that they do not harm each other significantly (…). But presently, we have no evidence of the mutual benefits that they could afford each other" (1878 [Oulhen et al. 2016: 136]). What de Bary finds particularly noticeable about the cases of lichenism are precisely the morphological and non-pathological effects of these types of associations, as well as the sorts of physiological dependencies that emerge from the partners. Drawing directly upon the experiments carried out by Stahl (1877), he highlights the important morphological changes that accompany the synthesis of lichens: "right after their association with the fungus of the lichen, the cells of the algae become much larger, contain more chlorophyll, [and] are stronger in every way. Beyond doubt, according to data, that have been known for a long time regarding the structure of the lichen, all of these characteristics are retained for the entire life cycle of the lichen, sometimes for several dozen years" (1878 [Oulhen et al. 2016: 138]). It is precisely at this place of his discussion, when de Bary applies Darwin's theory, explaining that symbiosis might work as an inducer of the morphological changes that are required for natural selection to generate adaptation (see also Sapp 1994: 9; Sapp 2003: 234–251).

What it is at stake in de Bary's discussion of lichenism is also a debate about the nature of biological individuality. If the elements that conform the lichen are dissociated, the lichen does not exist anymore, and we would only have a fungus – that will eventually die– and an alga. However, if we put them together to generate new lichen, they become somehow denaturalized, they lose their main morphological characteristics, adopting a new configuration that makes them different from their free-living counterparts. De Bary's insistence in the non-parasitic, non-mutualistic nature of lichens is also noteworthy. On the one hand, he seems to have identified a new dimension of the "living together", which was not reducible to van Beneden's categories. On the other hand, he still wanted to keep the concept of "symbiosis" to name the associations that, in his words "we can group under the term sociability" (1878 [Oulhen et al. 2016: 136]), including some associations that do not question the individuality of the partners involved (e.g. pollination). Jan Sapp argues that "[t]his was a strategic argument that was designed to ensure that lichens were not discarded as exceptions" (1994: 9). I agree with him, and I think de Bary actually believed he was identifying a very distinct phenomenon, which questioned the conventionally accepted boundaries of biological individuals.[7]

1.2 Roscoe Pound (1870–1964)

After de Bary delivered his lecture, research on symbiosis started growing and new cases were discovered: Karl Bradt discovered the presence of the symbiotic alga *Zooxanthella* in the bodies of *Hydra* and sponges (1881, in Sapp 1994: 11) and Patrick Geddes discovered the presence of non-pathogenic alga in sea anemones (1882); some time later, Albert B. Frank discovered the presence of fungi in the root of legumes, which, he hypothesized, was a symbiont with important physiological functions, naming it "mycorrhiza" (1885 [Frank 2005]). In this period, symbiosis practically became identified with mutualism to a point where the two terms became interchangeable, while the general meaning of which de Bary had suggested became lost (Sapp 1994: 18–34; cf. Martin and Schwab 2012, who argue that the association between symbiosis and mutualism lasted until 1970).[8]

It is precisely in this context that Roscoe Pound lectured his: "Symbiosis and mutualism," with the aim of disentangling the two concepts (Pound 1893). Pound started his paper by distinguishing three types of relationships between hosts and parasites: those where the host kills the parasite; those where the parasite kills the host; and those where "the host lives on side by side with the parasite indefinitely" and continues "[a] further development is attained in cases where the parasite and host not only live together, but are mutually beneficial, and, perhaps, even, in extreme cases, *inter-dependent*" (1893: 509; emphasis added). For him, following de Bary, symbiosis just meant "living together for a long time," and mutualism is just one of the forms that this

[7] It is important to note that lichenologists originally rejected Schwendener's dual hypothesis (e.g. Crombie 1886), denying in some cases the evidence, among other reasons because its acknowledgment would threaten "the hard-won autonomy of lichenists themselves" (Sapp 1994: 4), in so far as lichens would stop being an independent biological individual. Interestingly, lichenologists did not lose their autonomy and it was precisely the study of lichens as dual individuals that began challenging traditional ways of understanding biological individuals more generally. This is the first moment, to my knowledge, that the problem of symbiosis and the philosophical problem of biological individuality get engaged in a way that questions the traditional conception of what counts as a biological individual.

[8] One of the reasons why symbiosis became identified with mutualism during this period is related to the influence of the political ideas of the time, especially the anarchist ideas of Kropoptkin (1902). Readers interested in the influence of political ideas on symbiosis thinking can refer to Sapp (1994: 18-25) and Gontier (2016a).

"living together" might take. Yet, in most circumstances, he claims, this "living together" does not take the form of mutualism. Furthermore, he provides some clarification that is especially important for the debate about biological individuality: acknowledging that mutualism can take forms other than "living together," he says "it should be noted that the mutualism of which we are here speaking is mutualism of parasite and host –not mutualism of *independent* organisms" (1893: 509; emphasis added).Why that distinction between cases of *inter-dependent* organisms versus cases of *independent* organisms? What makes the former cases so special? We must recall that de Bary had explicitly said that he had no objection to using "symbiosis" to refer to those associations that can be grouped "under the term sociability." My impression is that Pound had already perceived the qualitative difference between those two types of association: whereas the latter do not compromise the concept of biological individuality, as the organisms that interact can be clearly recognized as independent, the former do, in so far as the organisms (1) live in close association during all their life cycle and (2) might become inter-dependent to such a agree as to form a new entity.

Granted, there is a qualitative difference between the two types of associations (i.e. sociability vs. symbiosis) in so far as the latter, but not the former, compromise currently accepted ideas about biological individuality. In particular, it challenges the idea that one individual can belong to only one species classified according to the criteria of systematists. The rest of Pound's paper is dedicated in arguing that the biologists of his time had tended to overemphasize the presence of mutualism, claiming to have identified it in many symbiotic associations, where it was not at all clear that the partners were acting mutually. First, he argues, mutualism does not occur in every lichen: it does not exist in homoeomerous lichens, or in what he calls "pseudo-heteromerous lichens," although evidence suggests that it might exist in heteromerous lichens, as they exhibit a complex interdependence among the fungus and the algae that form the lichen (Pound 1893: 511–513). Second, he analyses Frank's studies on mycorrhizas and, while recognizing part of Frank's discoveries, he refers to some evidence by R. Hartig, "a more sober and trustworthy writer than Frank" (1893: 516). He argued that:

> "Organisms are not given to gratuitously assisting one another. Mychorhiza [sic] undoubtedly exists (…). But that there is, in any of these cases, more than the ordinary symbiosis of parasite and host, has not been shown and is improbable. That every tree has its root system covered with mycelia, proves nothing. Every tree has its bark covered with lichens, its twigs with black fungi, and its leaves with parasitic fungi of every description." (1893: 516)

Finally, he considers the presence of *Rhizobium* in the root of *Leguminosae*.[9] He says that the evidence is uncertain, and although it might sometimes seem as if the *Rhizobium* were mutualists, "[t]he bacteria (…) are parasites. They are there for their own purposes, and are incidentally beneficial to the plant" (1893: 518). Moreover, while admitting that in some cases the symbiosis might lead to a mutualism –as the plants infected do better than those uninfected–he continues diminishing the evolutionary importance of these symbioses by criticizing some of Frank's observations:

> "To these probabilities, Frank adds certain characteristic improbabilities. (…) [T]hat the plant develops tubes or hyphae for the purpose of self-infection which it sends through its tissues. (…) [T]hat the roots of the *Leguminosae* possess the power of attracting *Rhizobia*, due, as he considers, to some secretion. This is too much for his followers, and I think all will agree that it is the last straw of an unsupportable load with which he has already burdened our credibility." (1893: 519)[10]

And he concludes his paper saying:

> "Ethically, there is nothing in the phenomena of symbiosis to justify the sentimentalism they have excited in certain writers. Practically, in some instances, symbiosis seems to result in mutual advantage. In all cases it results advantageously to one of the parties, and we can never be sure that the other would not have been nearly as well off, if left to itself." (1893: 520)

Even despite Pound's dismissal of the importance of mutualistic symbiosis, as well as its general importance, his example helpfully illustrates the general awareness of the phenomenon among biologists in the late nineteenth century. Especially remarkable is his insistence of distinguishing between those cases *where interdependence is generated* versus those *where two (or more) individuals can be recognized as different*. Second, and also remarkable, is his way of neglecting the individuality of the symbiotic aggregate. As he expresses here and there, even if in some *rare cases* the individuality of the symbiotic aggregate might occur, the organisms are there for their own

[9] Although he did not call them *Rhizobium*, but "tubercles," stating "For all that I have read and seen, I am satisfied that the parasites [in Leguminosae] are bacteria, and I see no reason for separating them from the rest of *Schizomycetes* as Schneider does. I even doubt the necessity of creating a separate genus for them, as Frank did in 1890, under the name of '*Rhizobium*'" (Pound 1893: 517).

[10] See Oldroyd (2013) to realize that some of Frank's observations were indeed true and Pound, while having a fair point about the lack of proper evidence for some of Frank's statements, could have not been more mistaken.

benefit, and many of them would probably be better outside the symbiosis.[11] These claims have two important consequences. First, it suggests that the general rejection of symbiosis research by biologists writing at this time was for the reason that it seemed to negatively affect the traditional conception of biological individuality and "struggle for life" (see also Sapp 2003, 2004). Second, it paves the way for a new and important conceptual change in symbiosis, the important division between symbiosis and other forms of sociality, forms that de Bary had considered as manifestations of the same phenomenon.

1.3 Albert Schneider (1863–1928)

The next important step in the development of the concept was because of Albert Schneider, who in "The phenomena of symbiosis" proposed a new understanding of the symbiosis as "*a continuous association of two or more morphologically distinct organisms, not of the same kind, resulting in a loss or acquisition of assimilated food-substances*" (1897: 925). There were three purposes to his paper: first, to distinguish clearly between cases of associations of living thing and cases of real symbiosis; second, to suggest the possible evolutionary origin of symbiosis, accounting for the default behaviour of organisms, which he understood as a "struggle for life"; and third, to classify different types of symbiotic associations. Of course, the three questions are closely connected to one another: once symbiosis is distinguished from mere "association," the classification of different types of symbiotic phenomena will be partially based on evolutionary criteria. Therefore, the different types of symbiotic relationships will be distinguished by degrees, from the forms that entail independent individuality of the organisms that interact, to those where the associated organisms lose their individuality and merge to form a higher level entity.[12]

To start with, Schneider begins by pointing out the fact that symbiosis is something "abnormal," as organisms will usually tend to compete with each other. Symbiosis, thus, requires long periods of time, phylogenetically, in which the organisms live in close proximity or real contact, and this makes it almost impossible to determine when the starting point of the symbiotic relationship occurred. After the relationship has begun, the organisms will start experiencing some "morpho-physiological" changes that will reinforce their connection, increasing their degree of mutual dependency. At some point, the two organisms will merge and their relationship will be one of holistic dependency with respect to each other; this is a phenomenon that Schneider calls "complete individualism." With this basic scheme of the evolution of symbiotic relationships in mind, Schneider classifies the different types of symbiosis as follows:

I. "Incipient Symbiosis (Indifferent Symbiosis)

1. Accidental Symbiosis
2. Contingent Symbiosis (*Raumparasitismus*)

II. Antagonistic Symbiosis

1. Mutual Antagonistic Symbiosis (Mutual Parasitism)
2. Antagonistic Symbiosis (Parasitism)

a. Obligative Antagonistic Symbiosis
b. Facultative Anatagonistic Symbiosis

3. Saprophytism

a. Facultative Saprophytism
b. Obligative Saprophytism

III. Mutualistic Symbiosis

1. Nutricism (Semi-Mutualistic Symbiosis)
2. Mutualism
3. Individualism

a. Semi-individualism
b. Complete individualism" (1897: 930–931)

Commenting on this classification Schneider makes two important remarks: first, the development of the different types of symbiotic associations and their particular character will depend largely on *environmental opportunity*; second, he puts the emphasis in studying "the phylogenetic relationship of the symbioses without any reference to the phylogeny of the organisms comprising them" (1897: 931). In this vein, as he indicates, it means that one does not need to study the phylogenetic evolution of the specific organisms that engage in the symbiotic relationship, but only the relative evolution of the physiological relationship. Figure 1, taken from Schneider (1897: 932) presents his phylogenetic schema of the physiological evolution of symbiotic relationships.

[11] Pound's seems to assume a concept of biological individuality similar to what Queller and Strassmann have recently called the "cooperation/conflict conception" of the biological individual (2009, 2016). For Pound, as it happens for the authors, symbiotic assemblages cannot be considered individuals in the proper sense, as the entities that engage in the symbiosis are in constant struggle with each other.

[12] It must be noted, although in passing, that Schneider does not require that the two organisms that engage in symbiosis belong to different species: he only requires that they are morphologically different. That's why, from his perspective, the mother and the embryo/foetus, the sexual cells that merge to form a zygote or even tumours or cysts would count as cases of symbiosis. This is, I think, different from de Bary's original purpose –probably that's why Schneider says that he uses symbiosis "in its broader meaning, not in the sense of De Bary" (1897: 923, fn. 1)–, who seemed to understand symbiosis requiring different species.

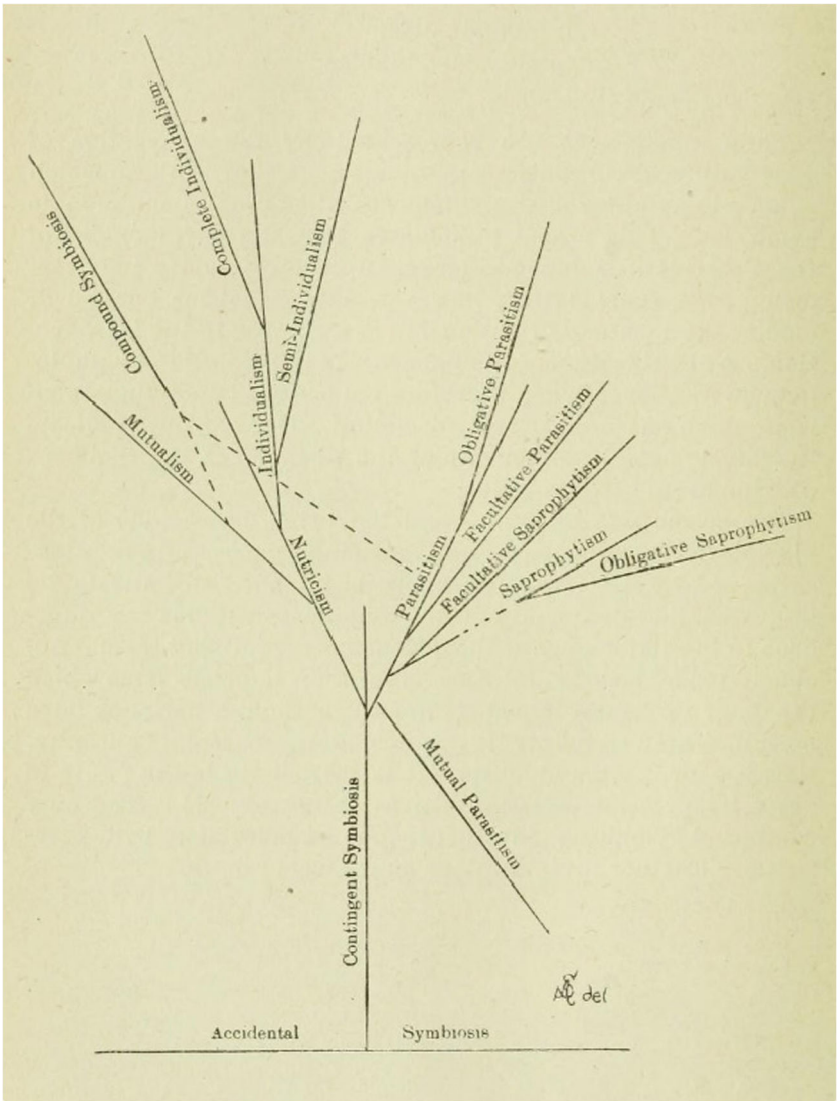

Fig. 1 Phylogentic development of symbiosis attending to physiological criteria. Although saprophytism is included in the figure, showing its relation to other symbiotic forms, it is not classified as a symbiotic relation, according to Schneider; rather, he provides it as a point of comparison

Schneider acknowledges from the beginning of his paper the difficulty of determining the starting point of the symbiosis. Under the heading of "incipient accidental symbiosis" he includes those cases where the organisms are in close physical contact for a sufficiently prolonged time, understood ontogenetically, and irrespectively of whether or not morpho-physiological changes (either antagonistic or mutualistic) occur.[13] Moreover, he argues that, once an accidental symbiosis has been established, then the condition will immediately be subject to change, since the permanency of this symbiosis is in direct proportion to the degree of mutualistic specialization (1897: 934). In this sense, if an accidental symbiosis is not broken down, it will evolve towards a "contingent symbiosis," where the organisms involved, despite not experiencing any morpho-physiological change, seem to manifest a sufficient degree of elective affinity. One case of contingent symbiosis, according to Schneider, is the bacterial flora of humans, which shows a certain degree of elective affinity but does not seem to show any kind of morpho-physiological relationship with the host.[14]

The second symbiotic phenomena considered are the cases of "Antagonistic Symbiosis." According to Schneider, this category includes "mutual parasitism," i.e. the situation where both organisms live together but their relationship is mutually damaging, "parasitism," a situation where one of the organisms is damaged whereas the other obtains benefit from the relation, and, as a limiting non-symbiotic case, "saprophytism". For Schneider, antagonistic forms of symbiosis can only give rise to very limited morpho-physiological specializations or

[13] Schneider acknowledges the problems of this position, which can be criticized on the same basis as Pound had criticized Frank's account of mycorrhiza –"[t]hat every tree has its root system covered with mycelia, proves nothing" (Pound 1893: 516). However, he justifies his decision by claiming "[f]rom a priori reasoning one is, however, forced to conclude that the first symbiotic activities began with the first contact of organisms" (1897: 933).

[14] See part II of the paper for seeing how these sorts of claims are presently unsustainable.

adaptation, since the parasitic nature of the relationship causes it to be "a destructive association, [such that] [t]he morphological and physiological changes tend towards dissolution rather than evolution" (1897: 936). Therefore, he has a reason to believe that, even if antagonistic forms of symbiosis are conceivable (and, in principle, even expectable), a case of antagonistic symbiosis either evolves towards a case of mutualistic symbiosis or it will be driven towards extinction.[15]

The final kind of symbiotic phenomena, the cases of "mutualistic symbiosis," can occur when two organisms interact with each other so that the relation is mutually beneficial. The mutual benefit might occur either because one organism benefits another without being damaged ("nutricism"), because both organisms "mutually benefit each other [while] are still capable of leading an *independent existence*" ("mutualism") (1897: 941), or because "*one* or *more* of the symbionts is absolutely dependent upon the other for its existence" ("individualism") (1897: 943). Schneider remarks, however, that it is very unlikely that something such as "absolute nutricism" really occurs in the biological world. He acknowledges that in some symbiotic associations one of the symbionts is clearly benefitted, whereas the material benefits for the other are not so clear. However, he thinks that in most cases nutricism will tend to evolve towards a relation of mutual benefit for both partners. This last type of relationships might happen either in cases where both symbionts can carry independent existence (he mentions insectivorous plants and their bacteria, *Actinia prehensa* and *Melia tessellata* or some species of ants and the branches of trees), or in cases where they are mutually dependent. About this last case he claims "[i]t (…) represents a higher form of mutualism, from which it is no doubt phylogenetically derived. (…) [In individualism] [t]he associations form an individual, a morphological unit, and the phenomena *are frequently not recognized as symbiosis*" (1897: 943, emphasis added).

It is important to realize, at this point, that Schneider's work is conceptually revolutionary. First, he is the first to consider the possibility of studying the phylogenetic history of symbiotic associations (i.e. their evolution): (1) irrespectively of the evolution of the organisms that form the symbiosis; and (2) relative to the opportunities that the environment offers for their evolution (in this sense, symbiotic assemblages would be something conceptually similar to what we now call "units of selection"). Second, he realizes that symbiotic associations challenge the individuality of the organisms that interact, to the point that they might become a new independent emergent individual. Conceptually speaking, Schneider is the first author to recognise this last fact, thus opening the possibility of understanding symbiotic associations as genuine evolutionary individuals in their own right. Furthermore, he is conscious of the physiological importance of symbiosis, as well as why it is occasionally not possible to understand the physiology of the organisms in isolation from their symbionts.[16] This fact has gained a lot of attention recently, especially after the hologenome concept of evolution was proposed.

1.4 Constantin Merezhkowsky (1855–1921)

A final and important step in the conceptual development of the association between symbiosis and biological individuality is because of the work of Constantin S. Merezhkowsky. In his "*Über Natur und Ursprung der Chromatophoren im Pflanzenreiche*," ("On the nature and origin of chromatophores [plastids/chloroplasts] in the plant kingdom"), Merezhkowsky proposed, for the first time, the term "symbiogenesis," further advancing the conception of symbiosis as an evolutionary mechanism (Merezhkowsky 1905, 1910) (see also Khakhina 1992; Sapp 1994: 47–59; Martin and Kowallik 1999; Sapp et al. 2002: 418–423; Gontier 2016b).

In his paper, Merezhkowsky aimed at discerning the origin of chloroplasts in plants. During his time, it was commonly believed that the chloroplasts, contained in the body of plants, appeared de novo every new generation and, furthermore, that they originated autogenously, as new organs which differentiated within the bodies of plant cells. Merezhkowsky strongly disagreed with that conception. Drawing upon Schimper's discovery that chloroplasts do not appear de novo in plant cells, but are always present within their bodies since the beginning of the life of the plant (1885, referred in Merezhkowsky 1905), he proposed a revolutionary notion: chloroplasts should not be regarded as autogenous organs of plants, but as symbionts, i.e. as independent (foreign) organisms that live together with plant cells. Merezhkowsky offered two different types of arguments to support his theory. His first two arguments were theoretical. The first one was based on Schimper's discovery: if chloroplasts do not arise *de novo*, though invaginations of the cytoplasm of the cell, but "rather, they always arise through division of pre-existing plastids, and since the latter in turn arise from pre-existing plastids, etc., we necessarily arrive at the logical conclusion that long ago the first chromatophore migrated into a colourless organism" (1905: 596 [Martin and Kowallik 1999: 289]).[17] Secondly, Merezhkowsky argued that chloroplasts can be understood by analogy to *Zooxanthella* in the body of *Amoeba viridis*. In both

[15] Of course, extinction *of the symbiotic association*, but not necessarily of the partners that interact symbiotically. Remember that Schneider's paper aims to study exclusively the phylogenetic evolution of symbiotic associations without reference to the organisms that interact.

[16] Those readers who are not familiar with the different types of biological individuals (physiological, anatomical, developmental, evolutionary, etc.) can check Gilbert et al. (2012), Godfrey-Smith (2013), Pradeu (2016a, b), DiFrisco (2017). In brief, however, it is important that she notes that not all criteria for classifying biological individuals necessarily led to coincidental classifications and sometimes different criteria overlap. For the overlapping nature of biological classification see Clarke (2010).

[17] It is widely acknowledged that chloroplasts are responsible for the green colour of plants.

cases, the structures (chloroplasts and *Zooxanthella*, respectively) can be said survive, divide and behave as independent organisms. If biologists do not have any issue in understanding *Zooxanthella* as independent symbiotic organisms within the bodies of their hosts, they should not have any prejudice in applying the same type of reasoning to chloroplasts, providing that the empirical evidence supported this conception.

The rest of the rationale to his hypothesis were of empirical observation. First, the discovery that chloroplasts, in contrast with other "organs" ("organelles") in the body of plant cells, can survive and reproduce even after the nucleus of the cell has been removed, and also can do so outside the cell's cytoplasm, which suggests that they behave like independent organisms. Second, the similarity between chloroplasts and free-living bacteria, concretely, with free-living forms of *Cyanophyceae*, which has been qualified by Martin & Kowallik as the "unquestionably most novel line of reasoning" (1999: 287). According to Merezhkowsky, chloroplasts and *Cyanophyceae* had a very similar physical appearance, both in form and colour, very similar biochemical (physiological) properties (with a similar type of nutrition), and analogous ways of proliferation and reproduction, which "makes it exceedingly likely that chromatophores are *Cyanophyceae* that invaded the plasma" (Merezhkowsky 1905: 600–601 [Martin and Kowallik 1999: 291]). Finally, he argued that, as it was empirically proven that *Cyanophyceae* can also engage in symbiotic relationships with other organisms (diatoms, rhizopods, etc.), even with cells that are protected by a cell wall, it was possible that at some point in their evolutionary history *Cyanophyceae* could have entered in contact with a plant cell so as to give rise to chloroplasts.

Merezhkowsky's symbiogenetic hypothesis, as well as his arguments, gives symbiotic ideas a new meaning. Authors writing prior to him had discussed the importance of the symbiotic relationship, the nature of the symbiotic relationship, how symbiotic relationships could cause several morpho-physiological changes in biological individuals, etc. However, no one had considered the possibility that symbiosis might be a hereditary phenomenon, i.e. that symbiotic associations might be intergenerationally transmitted (e.g. like gametes passing between germ-line cells). Authors had assumed that genetically heterogeneous organisms reproduced independently, and later would form symbioses. Merezhkowsky, on the contrary, challenged the necessity of this assumption; and, by implication, questioned the boundaries of biological individuals, understood evolutionarily. For instance, it is not just that different organisms engage symbiotically and later their morpho-physiological independence is lost; in the case of plant cells, also their hereditary independence (evolutionary individuality) is lost, as the two previously independent organisms are now inherited exclusively together. In summary, Merezhkowsky includes the main element that was lacking in the symbiosis picture, conceiving, for the first time, the idea of hereditary symbiosis. In one sense, symbiotic assemblages had already been attributed all the necessary elements for being considered units of selection (variance, inheritance, fitness; e.g. Lewontin 1970; Godfrey-Smith 2009). The question afterwards changed this sense, since it now asked us to determine the real importance of heritable symbiosis. Was this just an isolated case special to a different phenomenon or was it general to symbiosis as such?

The notion of hereditary symbiosis was later supported by Hermann Reinheimer, Andrei S. Famintsyn and Boris M. Kozo-Polyansky (1924/2010) (Khakhina 1992; Sapp 1994: 47–59; Carrapiço 2015). Afterwards, Paul J. Portier (1918) and Ivan E. Wallin (1927) would apply symbiogenetic ideas to the origin of mitochondria, extending Merezhkowsky's original application to another cellular organelle. And even later, Paul Buchner would explore the importance of hereditary symbiosis in insects, proposing a new field of application for the hypothesis (Boucher 1965; Sapp 2002). Symbiogenetic theories of the origin of the eukaryotic cell, however, were frequently rejected. This trend continued for almost 50 years, until Lynn Margulis provided new support and the symbiotic origin of the eukaryotic cell became almost universally accepted (Sagan 1967; Margulis 1970, 1991, 1993; see Sapp 2010). Nonetheless, it is important to remark that the conceptual basis for understanding the role of symbiosis in evolution, as well as the possibility of considering some symbiotic assemblages as what we would call "units of selection" in contemporary jargon, were already settled by Merezhkowsky in 1905. By then, all the conceptual connections between the notions of symbiosis and biological individuality were already present, as well as the conceptual challenges that the former presented for traditional conceptions of the later. In the next part of the paper, I will explore how those conceptual connections have been explored in recent times, especially after the proposal of the hologenome concept of evolution.

2 Part II. Holobionts and hologenomes – contemporary philosophical implications of symbiosis

The previous part of the review has analysed how the concept of symbiosis appeared in biology and how the connections between symbiosis and biological individuality changed and developed. Towards the end of this history, the concept would express itself, in the form of hereditary symbiosis or symbiogenesis, and some biologists postulated it as a mechanism of evolution. This part of the review will examine recent conceptual debates in the symbiosis literature, especially the notion of the holobiont and the hologonome concept of evolution. In section 1, I discuss Lynn Margulis' introduction of the concept, as well as its relation to Merezhkowsky's notion of "symbiogenesis". I argue that Margulis' view of the holobiont is ambiguous: sometimes the holobiont is

apparently restricted to cases of hereditary symbiosis and other times it is not. In section 2, I discuss the hologenome concept of evolution, as Rosenberg and Zilber-Rosenberg introduced it, and I review the current debates that it raises in connection to the problems presented by the concept of biological individuality. In section 3, I relate the hologenome concept of evolution to the historical discussion presented in Part I, arguing that the conceptual disputes that the hologenome concept has generated are not new, but only a progression of the previous disputes that were held in the nineteenth century. I observe four parallelisms which obtain between the past disputes on symbiosis and the disputes raised by hologenomes, as well as three further points of distinction.

2.1 The origin of the concept – the importance of Lynn Margulis (1938–2011)

Lynn Margulis (born Alexander) was a pioneer in the field of symbiosis, to which she dedicated almost 50 years. She is especially known for giving new life to the hypothesis of the symbiotic origin of eukaryotic cells, as well as for her enthusiasm about the importance of symbiosis for life on Earth and evolution (Margulis 1990, 1991, 1998, 2010; Sagan & Margulis 2002; Díaz 2015; O'Malley 2017). Margulis is acknowledged as the first person to introduce the term "holobiont", which was published in her paper "Words as battle cries – symbiogenesis and the new field of endocytobiology" (1990). In this work, she compares cyclical hereditary symbiosis with meiotic sex. In both of these compared cases of inheritance, she argues that two entities are present, which cyclically recognize each other and merge together for every generation. Moreover, in both of these cases she speculates the presence of mechanisms, which guarantee the integration of these two entities and, also, their subsequent dissociation, resulting in the formation of a new individual. An entity formed of two different gametes is what we call a "zygote," whereas the entity that results from the merger of two symbionts is what Margulis refers to as "holobiont", which she recognises as a new individual (1990: 676, Fig. 3). Margulis does not, however, specify which "bionts" should be regarded as part of the holobiont, nor does she explicitly define the term in the paper.

One year later, in "Symbiogenesis and symbioticism", the first chapter of a book she edited with René Fester, Margulis defines the holobiont as a "symbiont composed of recognizable bionts", and she defines symbiosis as the physical contact between organisms of different species occurring "throughout a significant proportion of the life history" (1991: 2, Table 1). Again, she does not explicitly specify which bionts should be included in the holobiont. If one follows her definition of life history strictly –"events throughout the development of an individual organism correlating environment with changes in external morphology, formation of propagules, and other observable aspects" (1991: 2, Table 1)– it might be argued that the holobiont would encompass all the bionts that share their lifetime together, irrespective of whether they are inherited or not. Clearly, this conception of the holobiont would be incoherent with the concept she had put forward in her previous (1990), where she seemed to suggest that the holobiont should exclusively include the cases of hereditary symbiosis, in her analogy between symbiogenesis and embryogenesis. This second formulation is reasonable if one takes into account the purpose of the chapter, namely, to vindicate the proposition of symbiogenesis as a way in which new species, kingdoms and taxa could evolve –for instance, she says that "the highest level taxa (…) have evolved by acquisitions of symbionts *that have become hereditary*" (1991: 11, emphasis added)–. This formulation is also coherent with claims she made in her later writings (Margulis and Fester 1991; Margulis 1998, 2010; Margulis & Sagan 2001, Margulis and Sagan 2002; and also see O'Malley 2017). For instance, in one of her latest paper, where she justifies the historical role of Kozo-Polyansky in introducing the idea of symbiogenesis to biology, she argues for the necessity of genetically distinct bionts reproducing together in order for symbiogenesis to occur. Analysing the association between eels and a specific species of shrimp (cleaning symbiosis), she argues:

> "It is symbiosis, but not symbiogenesis. Both partners *grow and reproduce separately*. Both shrimp and eel can live separately. One sees no obvious novelty generated by this symbiosis; i.e., symbiotic physical association. The relationship between the shrimp and the eel is still a behavioral one" (2010: 1528, emphasis added)

In this vein, one might argue that, as "holobiont" was introduced in comparison to meiotic reproduction, and Margulis discusses it while reflecting the importance of symbiogenesis as an evolutionary mechanism (and evolution requires inheritance), the holobiont is thus the biological individual that includes all those symbionts that are inherited together (organelles in eukaryotes, obligatory endosymbionts in insects, etc.) (O'Malley 2017: 36, for a defence of this interpretation).

This interpretation of Margulis' understanding of holobionts is not without contestation, though. In the same volume where Margulis published her paper, Maynard-Smith suggests "a Darwinian view of symbiosis" (Maynard-Smith 1991). There, he relates the problem of symbiosis to the problem of the units of selection[18] and embeds it in the framework of the theory of

[18] Maynard-Smith does not use "units of selection", but "units of evolution", where a unit of selection is whatever entity exhibit phenotypic variation that led to multiplication of the entity within the population (thus being selected for or against), and a unit of evolution is a unit of selection that, furthermore, exhibits heredity (Maynard-Smith 1987). In contrast with Maynard-Smith, I will use "unit of selection" as it is conventionally used, i.e. requiring heredity, variance and fitness/multiplication, and thus meaning what Maynard-Smith means by "unit of evolution" (see Lloyd 2017a, c: 293–297; Gontier 2010, for an analysis of the concept of "unit of selection")

evolutionary transitions in individuality that he was starting to develop. According to Maynard-Smith, symbiosis can be understood as an evolutionary mechanism and interpreted in a Darwinian fashion (i.e. with the entities that interact symbiotically being a unit of selection) only if the entities that interact symbiotically are transmitted directly, because "[w]ith direct transmission the genes of the symbionts will leave descendants only to the extent that the host survives and reproduces" (1991: 35). Therefore, as far as the two bionts have their fitness interests aligned, it is expected that those symbionts will tend to maintain a mutualistic relation that, eventually, might make it "reasonable to consider the association as a single unit" (1991: 38). However, in cases of indirect transmission, this possibility is much less likely, and he suggests that the interacting entities should be considered as independent units (of selection).

Maynard-Smith's paper is relevant because he seems to be discussing Margulis' liberal views about the power of symbiosis. For him, those cases where symbiosis might be considered to have evolutionary power, in the sense of affecting the role of natural selection, are very limited, and probably precluded only to cases such as cellular organelles, as he suggests at the end of his paper. If this is so, then Margulis' notion of the holobiont might be interpreted not as constrained exclusively to the cases of the eukaryotic cell, but as including the associations of many different bionts. In fact, this view is endorsed in Guerrero et al. (2013), published two years after Margulis' death. In that paper, holobionts, considered as autopoietic (self-sustaining) units, are defined as "integrated biont organisms, i.e., animals or plants, *with all of their associated microbiota*" (2013: 133, emphasis added). In the same place, they also coined the term "holobiome", referring to "the assembly of genetic information contributed by the animal or plant and its associated microbiota" (2013: 134), and demanding a new look at evolution that would take into account the importance of the host genome plus the genome of its microbiota. They argued this to be a new entity, whose basic interacting elements that would give rise to new species and, in general, new biological variety. At some point of the paper, the authors even endorse the theses that: (1) holobionts are subjected to natural selection; and (2) holobiomes are entities that have been selected due to their selective advantages. Even if the authors do not mention the concept "units of selection", their paper might be interpreted as endorsing the hologenome concept of evolution, thus considering the holobiont, with its hologenome (holobiome), as a possible unit of selection in evolution.

Whether Margulis' concept of the holobiont has to be interpreted as encompassing only hereditary symbiosis or, on the contrary, encompassing the whole collection of symbionts, and whether she was claiming that holobionts are units of selection or not, it seems clear that her conceptual heritage in the field of symbiosis is very important. She was one of the most vigorous defenders of the role of symbiosis for causing novelty in evolution (Margulis 1998; Margulis and Sagan 2002). Moreover, she coined the notion of the "holobiont", which is one of the most discussed concepts in philosophy of biology at present. In the next section, I analyse the recent usage of the notion of the holobiont, as well as the criticisms that have been raised against it.

2.2 The hologenome concept of evolution and its critics: a review of current debates

The hologenome concept of evolution[19] was originally proposed by Eugene Rosenberg and collaborators (Rosenberg et al. 2007), in their review paper: "The role of microorganisms in coral health, disease, and evolution", as a generalization of the coral probiotic hypothesis (see Reshef et al. 2006).[20] Drawing upon their observations on coral disease, the authors suggested the existence of: "a dynamic relationship (…) between symbiotic microorganisms and corals at different environmental conditions that selects for the most advantageous coral holobiont in the context of the prevailing conditions. By altering the structure of its resident microbial community, the holobiont can adapt to changing environmental conditions more rapidly and with greater versatility than a process that is dependent on genetic mutation and selection of the coral host" (2007: 360).

Moreover, reasoning from the existence of this dynamic relation between the coral host and its microbiota, as well as the knowledge that the possibility such a relation offers for the adaptive evolution of a coral to changing environmental conditions, the authors inferred that the coral holobiont must be a unit of selection, i.e. that it is subjected to the process of evolution by natural selection. Drawing upon the observation that, as it happens in corals, all animals and plants harbour an

[19] Originally, they referred to it as the hologenome theory of evolution. Later on, they started calling it the hologenome concept of evolution (cf. Gissis et al. 2017: 303–384).

[20] A clear antecedent to the hologenome concept is found in Sapp (2003: 234-251, 2004), when he coins the concept of "symbiome". He defines the symbiome as the entity "comprising chromosomal genes, organellar genes, viral genes, as well as other microbial symbionts, sometimes inside cells and always outside them, functioning across a continuum from parasitism to mutualism, depending on their nature and context (…). Since every plant and animal consists of complex ecological communities of microbes, the symbiome must function as a unit of selection." (2004: 1047). Nonetheless, Sapp first presents the concept in a section dedicated to developmental symbiosis (Sapp 2003: 235–236), and there is no reason to believe that a developmental organism should be delineated by the same boundaries than a unit of selection (e.g. DiFrisco 2017). The concept of "symbiome", however, is not as frequent in current literature as the concept of "holobiont" and it has been recently used with two different meanings: first, to refer to the whole set of symbionts that associate with a host, without including the host (e.g. Boucias et al. 2013; Rosas-Pérez et al. 2017); second, to refer exclusively to "the colocalized and coevolving taxa in a given consortium" (Tripp et al. 2017: 552). If we define the concept according to the second formulation, then one might argue either that *symbiome = hologenome* (if the hologenome is proven to evolve as a single unit) or that the symbiome corresponds to the part of the hologenome that actually evolves as a single unit (e.g. the set of vertically transmitted symbionts). This warrants further discussion, which is, however, outside the scope of this paper. For my present purposes I will restrict the discussion to the concept of the holobiont *sui generis*.

abundant number of symbiotic microorganisms in their bodies, the authors suggest that we generalise the coral probiotic hypothesis to include every animal and plant. Thus, Reshef et al. proposed the hologenome concept of evolution, the notion that "the holobiont with its hologenome should be considered as the unit of natural selection in evolution, and microbial symbionts have an important role in adaptation and evolution in higher organisms" (2007: 360, Box 2).

Nevertheless, in the original paper, the authors do not specify: the meaning of "holobiont", the meaning of "hologenome", or how their hypothesis could be applied to other model organisms. Instead they briefly justify its appeal on four grounds: first, the universality of symbiosis between animals/plants and microorganisms; second, the existence of phenotypic variance between host species and their microbiota, i.e. the fact that hosts of the same species harbour different microbiotas; third, the different range of effects of the microorganisms on their hosts (parasitism, mutualism, commensalism); and fourth, the possible mechanisms of change for the holobiont (including microbial amplification, microbial acquisition, etc.) (Rosenberg et al. 2007: 360, Box 2). However, the authors acknowledged that their reasons were insufficient to support their generalization of the coral probiotic hypothesis. To overcome this difficulty Zilber-Rosenberg and Rosenberg (2008) would publish "Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution" one year later. Beginning with the acknowledgment that microorganisms have been discovered to play a fundamental role in the life of higher organisms (animals, plants), including humans, the authors introduced their hypothesis with a rhetorical question:. "[i]f microbial symbionts play such an important role in the lives of their eukaryotic hosts, why should they not also play a role *in the evolution* of these higher organisms?" (2008: 723, emphasis added). Zilber-Resenberg and Rosenberg hypothesized that holobionts (i.e. biological entities composed by a host plus all its microbial symbionts), with their hologenomes (i.e. the sum of all the genetic information of the host plus the genetic information of its symbionts) are units of selection. More specifically, concerning the notion of the holobiont, they explicated that:

> "Although much of the important research on symbiosis has been carried out with a small number of model systems involving a single major symbiont, the hologenome theory places importance not only on these major symbionts but also on the enormously diverse associated microbiota, which have only been uncovered in recent years using molecular techniques" (2008: 724)

This last point is particularly relevant because it frames the hologenome concept in a very distinctive way. It is not just that very particular host-microbe associations should be considered as units of selection (e.g. the eukaryotic cell, aphids and *Buchnera aphidicola*, squids and *Vibrio fischeri*, etc.). This last proposal would not be so revolutionary, after all. The hologenome concept suggests that one should consider the host, *with all its microbes* (i.e. the holobiont), as a unit of selection in evolution. Notice that this definition of the holobiont might be contrasted with Margulis' understanding, which seemed to be limited to cases of hereditary symbiosis, at least according to some interpretaters (e.g. O'Malley 2017). What is the justification that Zilber-Rosenberg and Rosenberg believe to have found for their hypothesis? They claim the existence of four sources of evidence: the observation that all higher organisms associate with microorganisms; the fact that symbionts are reliably transmitted intergenerationally; the fact that symbionts affect the fitness of the holobiont; and, finally, the possibility of generating genotypic variation within the holobiont by changing their microbial composition.

It must be noted that the way in which Rosenberg and Zilber-Rosenberg present the hologenome concept is based on a particular interpretation of the units of selection, according to which two types of questions should be distinguished: first, the question about the interactor, or vehicle, the entity that interacts with the environment as a cohesive whole, in such a way that replication is differential[21]; second, the question about the replicator, the entity of which copies are made (Dawkins 1976; Hull 1980; Okasha 2006; Godfrey-Smith 2009; Lloyd 2017a). For Zilber-Rosenberg and Rosenberg, the holobiont would be an interactor, a cohesive physiological and metabolic entity, whereas the hologenome would be a replicator (see also Rosenberg et al. 2010; Rosenberg and Zilber-Rosenberg 2014, 2016; Author 2015; Bordenstein and Theis 2015; Shropshire and Bordenstein 2016; Theis et al. 2016).

After Rosenberg and Zilber-Rosenberg proposed their hypothesis, the notion that the holobiont with its hologenome constitutes a biological individual has been defended in different ways by different authors, some of which have interpreted it as a unit of selection. Dupré and O'Malley (2009), and John Dupré (2010, 2012) have defended the notion that the holobiont should be considered as the interactor in evolution, in so far as it is the entity responsible for the differential reproduction of the entities that compose it. The authors do not mention, however, the possibility of conceiving the hologenome as a replicator. Scott F. Gilbert, Jan Sapp and Alfred I. Tauber have suggested that we understand the holobiont as a biological individual anatomically, developmentally, immunologically, physiologically and genetically

[21] "Such that replication is differential" does not specify which are the entities whose differential replication might be affected by belonging to an interactor. It is conceptually possible that the holobiont is an interactor that promotes a more efficient replication of the different individuals that compose the holobiont (host, microbes of the microbiome), but not of the hologenome.

(Gilbert et al. 2012; also Gilbert et al. 2017; Roughgarden et al. 2017). Lynn Chiu and James Griesemer have separately proposed a concept of the holobiont as a developmental hybrid in which the microbes would act as scaffolds of the individuality of the host (Gilbert & Chiu 2015; Chiu and Eberl 2016; Griesemer 2016, 2017). Lisa Lloyd has suggested an understanding of the holobiont as an interactor, as a reproducer, and as a manifestor of adaptation (Lloyd 2017b; see also Griesemer 2017). Ford Doolittle and Austin Booth have proposed to conceive the hologenome as a functional replicator, i.e. as a network of genetic interaction patterns that can be instantiated across different generations of holobionts (Doolittle and Booth 2017; see also Lemanceau et al. 2017); Suárez (under review) has defended a group-selection interpretation of the holobiont, suggesting that we conceive of holobionts as intergenerationally inherited collections of traits associated to successive generations of a particular host. In so far as holobionts can be considered collections of traits, he argues that they can be conceived of as units of selection. Finally, Ehud Lamm has suggested that holobionts should be understood as "structures of evolution": "*constellation[s] of evolutionary factors and their relations* […] [that] provide scientists with a common framework and terminology and [allow them] to elicit research questions and hypothesis that apply to many systems of interest" (2017: 372).

Furthermore, some evidence has been gathered in support of the hologenome hypothesis (e.g. Rosenberg and Zilber-Rosenberg 2014; Bosch & Miller 2016, for general summaries). In a pioneer study on *Nasonia* wasps, Robert M. Brucker and Seth R. Bordenstein have argued that hybrid lethality among different *Nasonia* species is caused by a disruption of the relation between their species-specific microbiomes and the host genome, which suggests that the different species represent a coevolved hologenome (2013; cf. Chandler & Turelli 2014 for a response; cf. Brucker & Bordenstein 2014).Their study has prompted an immediate interest in the study of the phenomenon of phylosymbiosis, "the eco-evolutionary pattern, whereby the ecological relatedness of host-associated microbial communities parallels the phylogeny of related host species" (Brooks et al. 2016: 1). Convergent host-microbe phylogenies that support the existence of phylosymbiosis have been found in hominids (Ochman et al. 2010; Moeller et al. 2016). Julia K. Goodrich and collaborators have found some evidence that suggests that the microbiome might be heritable and its composition could be partially determined by the host genome (Goodrich et al. 2014, 2016, 2017; see also Turpin et al. 2016). Finally, Thomas W. Cullen and collaborators have found some evidence that might suggest that the host's immune system might control microbiota acquisition (Cullen et al. 2015). However, some evidence has also been found that suggests that there are no such tight host-microbiome intergenerational associations. For instance, Eric R. Hester and collaborators have not found evidence that supports inheritance of the microbiome among corals. Instead, they found that the microbiota that associates with a coral species are selected according to functional criteria, and thus there are no intergenerational phylogenetic convergences (Hester et al. 2016). The same results have been found in ruminal ecosystems: even if the hosts of the same species might share a functionally similar microbiota, the specific microbial taxa that they associate with are different. The authors explained the occurrence of this phenomenon with a metaphor: "the players might change but the game remains" (Taxis et al. 2015; Doolittle and Booth 2017 base their account of the holobiont on these results).

The hologenome concept, however, has also been contested by many, who propose that: (1) the holobiont is a sufficiently coherent biological entity for it to be considered an evolutionary interactor (Booth 2014; Queller and Strassmann 2016; Skillings 2016); there is no real empirical evidence supporting the claim that the hologenome can be a replicator or a reproducer, in so far as the fidelity of its intergenerational transmission is very low (Moran and Sloan 2015; Godfrey-Smith 2015; Stencel 2016; Douglas and Werren 2016; Hester et al. 2016; Hurst 2017; Stencel & Wloch-Salamon under review). Detractors of the holobiont concept tend to emphasize the lack of shared interests and unifying mechanisms between the entities that compose holobionts; and, on this basis, they are reluctant to accept the notion that holobionts are units of selection in any of the aforementioned senses.

The claim that holobionts are interactors has been recently disputed by Austin Booth who, emphasizing the fact that the different entities that compose a holobiont can reproduce independently, has argued that "the interactor perspective on holobionts, as currently endorsed, suffers from imprecision. More needs to be said about just what kinds of causal interactions among parts serve to bind independently reproducing populations into interactors" (2014: 670). This notion has also been criticized by David C. Queller and Joan E. Strassmann, who argued that holobiont defenders make an illegitimate inference from physical proximity (symbionts living together) to functional integration (symbionts constituting an interactor): "The holobiont is defined by spatial criteria. There is no reason to believe that spatial proximity necessarily leads to functional integration" (2016: 869). And also Derek J. Skillings has criticized this notion on the basis that the entities that compose the microbiome of a holobiont might change during the host's lifetime. If this is so, he argues, then there are no criteria of identity to recognize a holobiont as a biological individual (sensu organism or interactor), because the microbial species that compose it are constantly and fluidly changing (Skillings 2016).

In relation to the claim that holobionts are replicators, Angela E. Douglas and John H. Werren have rejected the possibility on the basis that holobionts lack the proper type

of intergenerational inheritance (Douglas and Werren 2016). For them, the holobiont can be considered a unit of selection if and only if there is sufficient partner fidelity –"stable association of host and symbiont genotypes across multiple generations" (2016: 2) – among the different species that constitute the holobiont. Otherwise, the entities that compose the holobiont would not have their fitness interests aligned; and thus, selection at the level of the holobiont would be disrupted by selection at lower levels. They concede that very specific and tight host-symbiont associations, under very special circumstances, may qualify as units of selection. However, they are sceptical that the same might be said about *all* the members of the microbiota: "We do not argue that selection cannot act on the host-microbiome as a unit. We simply argue that the evidence for this is weak, and the conditions necessary for it to occur are unlikely" (Douglas and Werren 2016: 5; see also Moran and Sloan 2015; Hurst 2017). Suárez (under review) has offered a specific reply to this criticism, arguing that their requirement of partner fidelity is unreasonable, since it relies on some assumptions about biological individuality that are disputable (the cooperation/conflict concept of biological individuality). Furthermore, he argues that the same type of assumptions are not applied to other levels of the biological hierarchy (e.g. transitions in evolutionary individuality), which creates a disparity of criteria. Finally, Peter Godfrey-Smith has also criticized the notion that holobionts are reproducers on the grounds of his concept of Darwinian populations (2015). He believes that host-microbe associations can only qualify as units of selection in the situations when the host is able to "kidnap" the reproduction of the microbe, i.e. when host and microbe can only reproduce together as a unit, but not independently from each other, since otherwise the system would be disrupted. He claims this to be true of eukaryotic cells generally. Godfrey-Smith also acknowledges the existence of intermediate reproductive stages (i.e. reproduction partially kidnapped, but with a high degree of independence). In any case, he does not believe that there is any evidence to qualify the holobiont, conceived as the host plus all its microbes, as a unit of selection, because the parts can still reproduce independently of the whole and thus will not have the same interests.

The debates between defenders of the hologenome concept of evolution and its detractors reflect diverging conceptions of biological individuality. Defenders of the hologenome concept tend to emphasize the collaborative nature of life, as well as the importance of symbiotic associations for maintaining life as we know it. They seem to share a commitment to a view of biological individuality according to which the existence of conflicts amongst the parts of a system does not rule out the possibility of the system evolving as a unit. Furthermore, they concentrate on studying symbiosis as an independent phenomenon, and try to understand the evolution of symbiotic relationships by partially abstracting away from the organisms that engage in symbiosis. Detractors of the concept, on the other hand, tend to emphasize the impossibility of having a biological individual if the parts of the systems are in conflict with one another, thus rejecting any claim about the individuality of holobionts. They are prone to consider holobionts as mere ecological communities of independent organisms that are together due to environmental convenience, not due to shared evolution. They put more emphasis on the study of the different species that engage in the symbiosis that in the study of the evolution of the symbiotic relationship itself. More research is needed to determine the empirical consequences of the hologenome concept of evolution, as well as to unravel the empirical consequences of the different conceptual assumptions made by defenders and detractors of the notion. Research on the historical roots on some of the recent debates will help to determine the origins of some of the present assumptions in current debates, as well as help with clarifying different issues raised by the hologenome concept of evolution, some of which were already present in the debates of prior literature.

2.3 The historical roots of the hologenome concept of evolution

Most of the debates about the hologenome concept of evolution explored in the previous section parallel some of the debates about symbiosis explored in Part I. I will explore four parallelisms between them, uncovering the similarities between recent research and the research conducted in the nineteenth century. Finally, I will explore the novelties introduced by the hologenome concept of evolution, exploring its differences to previous research.

2.3.1 First parallelism. The importance of mutual dependence among organisms

One of the aspects of life that symbiosis research has emphasized since it originally appeared was the mutual dependencies that exist among organisms. Without being necessarily mutualists, organisms frequently rely and depend on each other in order to survive and reproduce. De Bary found that some of those dependencies were not just circumstantial, but were maintained throughout the entire life cycle of certain individuals of different species. After de Bary, many other scientists stressed the importance of mutual dependencies among organisms in order to sustain life as we know it. Defenders of the holobiont concept stress the existence of those mutual dependencies among organisms, putting a special emphasis on the interactions of animals and plants with their microorganisms. They frequently insist that the phenotypes of known animals are not the result of a genetic plan that develops without external influences, but are the result of a co-construction between the animal themselves and their symbionts. They stress that

anatomically, immunologically, developmentally and physiologically we have never been individuals, if individuality is understood as the linear development of a single genetic plan (Dupré 2010; Sapp et al. 2012).

Like the original proponents of the symbiotic concept, including de Bary, Schneider, and Merezhkowsky, contemporary defenders of the holobiont pay special attention to those processes of co-construction and criticize previous approaches that have tended to diminish the importance of collaborations for essential processes. From a conceptual point of view, defenders of the holobiont are not proposing anything new: the founders of symbiosis research knew already that the long-term ontogenetic interactions of individuals of different species questioned basic ideas about the defining boundaries of the individual. If a biological individual is a functional whole that can survive by its own means to a great extent, then individuals do not necessarily match species, as there could be individuals that are composed by two or more different species that interact to form a cohesive bigger whole (see ft. 5). Therefore, defenders of the hologenome share their ideas with the founders of symbiotic thought; and, in this sense, their theses are not conceptually new.

2.3.2 Second parallelism. Spatial continuity and biological individuality

One of the arguments used by the detractors of the hologenome concept was based on the fact that from the observation that two entities live together one cannot infer that these two entities are a biological individual. This criticism was explicitly expressed by Queller & Strassmann, who denounced the defenders of the holobiont who inferred functional integration of the entities that compose the holobiont from the fact that they live in physical contact (2016: 819). A similar view is expressed by Booth, when he demands the presence of concrete mechanisms that guarantee that the members of holobionts are genuinely bounded together. It is not enough to say that they reside within the body of the host. The type of relationship that those microorganisms maintain with their host needs to be specified, or otherwise their common boundedness would be biologically irrelevant.

Queller & Strassmann's and Booth's observations match closely with the observation made by Pound in 1893. Criticizing Frank's assumptions about the symbiotic character of mycorrhizas, he argued "[t]hat every tree has its root system covered with mycelia, proves nothing. Every tree has its bark covered with lichens, its twigs with black fungi, and its leaves with parasitic fungi of every description." (1893: 516). His criticism, even if it was directed at a different type of association, rests on the same kind of assumptions about biological individuality. To prove that two entities living together are a biological individual, one needs to prove that there is a shared functionality. Inferring that two entities are a unique individual (or that they relate to each other mutualistically) from the fact that they share the same physical boundaries is insufficient. Therefore, the criticism raised by Queller & Strassmann and by Booth cannot be considered as conceptually novel. It is true that the criticism applies to an entity that, intuitively (i.e. based on physical appearance), might be considered more "individualistic" than the association between mycorrhiza and trees, which Pound discussed. However, this does not mean that the structure of the arguments used to criticise the concept are different.

2.3.3 Third parallelism. Studying the symbiotic phenomenon independently of the organisms that engage in the symbiotic relation

Defenders of the hologenome concept tend to emphasize the functional relations that exist between specific hosts and their microbiota. Different researchers have stressed the importance of a proper and balanced microbiota for the healthy physiology (and development) of organisms. From this observation, many authors have inferred evolutionary consequences, as well as a history of shared coevolution among independent genomes that form a hologenome. In some cases, like in Doolittle and Booth's (2017), the hologenome has been defined functionally, as a set of functionally relevant genetic networks that are reconstructed again and again in every new realisation of a holobiotic unit. This functional view of the holobiont and the hologenome abstracts away from the organisms that interact symbiotically. What matters is that the same functional relationships reoccur every generation, as well as the evolution of those relationships, irrespectively of the organisms that guarantee that this happens.[22] This position contrasts with organism-centred views of symbiosis, in which what is significant is not so much the evolution of symbiosis itself, but the evolution of the organisms that engage in the symbiotic relationship.

Schneider might be taken as a key reference for those positions, in so far as his work emphasized the study of the evolution of the symbiotic phenomenon in itself, irrespectively of the organisms that engage in the symbiotic relationships. As he argued, symbiosis research should study "the phylogenetic relationship of the symbioses without any reference to the phylogeny of the organisms comprising them" (1897: 931). Furthermore, he also emphasized the importance of environmental opportunity for establishing symbiosis. The new functional approach towards understanding symbiosis could be conceptually understood in the terms of Schneider, and it seems significantly connected to his prescriptions about how to study the phenomenon of symbiosis and the different symbiotic relations that exist in nature. In this sense, conceptually speaking, the emphasis on physiology for

[22] This position is taken to the extreme in Doolittle (2017).

understanding the nature and evolution of symbiosis is not new at all; it was already present in past literature.

2.3.4 Fourth parallelism. From symbiosis to symbiogenesis. The origin of new individuals through symbiosis

Merezhkowsky famously emphasized the importance of symbiosis as an evolutionary agent that can generate evolutionary novelties (new structures) as well as new biological individuals. In this vein, he was situating symbiosis outside of the realm of ecology and putting it in the realm of evolution. Even if the importance of symbiosis for evolution had been also emphasized for other authors (de Bary, for instance, who pointed out the possibility of using symbiosis for doing evolutionary experiments and emphasized the importance of symbiosis in creating new biological structures), it was Merezhkowsky who first appreciated the possibility of generating new biological individuals as a consequence of the symbiotic merger of two previously extant ones. Defenders of the hologenome concept have exploited this last possibility and applied Merezhkowsky's ideas generally, not only to eukaryotic cells. If mitochondria are former symbionts that are now considered parts of a new biological individual (i.e. the eukaryotic cell), so too should the microorganisms that compose an animal's microbiota be considered parts of a new individual.

Conceptually speaking, there is no big difference between the hologenome concept and Merezhkowsky's ideas about symbiogenesis. In both cases, it is assumed that new individuals can emerge through symbiosis and that these new individuals will have new biological properties. Furthermore, in both cases symbiosis *qua* symbiogenesis is put in the realm of evolution, and is not considered exclusive to the realm of ecology. In this sense, the hologenome concept is not conceptually revolutionary, as the ideas were already present in early twentieth century biology. The qualities that differentiate the hologenome concept from previous developments in symbiotic thought, thus, must lie elsewhere.

2.3.5 What is new about the hologenome concept?

Despite the parallelisms between the hologenome concept of evolution and previous issues treated in symbiosis research, it cannot be denied that the hologenome concept of evolution has brought new interestto some debates in biology and philosophy of biology. Furthermore, it has introduced an element of novelty in those debates; this is the reason why the hologenome concept of evolution should be explored further, especially in order to find what precisely makes it unique. I think that there are three important elements that distinguish the hologenome approach from its historical antecedents:

– First, the hologenome concept of evolution appeared after the "omics" revolution, a moment when the technological tools available for scientific research allowed biologists to discover an important number of microorganisms that had been previously unnoticed (Rosenberg and Zilber-Rosenberg 2014). In this sense, and in contrast with previous research on symbiosis, the hologenome concept is more universal, as it departs from the empirical evidence that all animals and plants bear an important number of microorganisms within their bodies. Previous research on symbiosis, however, had not been able to detect the universality of the phenomenon, and only some specific cases of symbiosis were studied. In addition, in previous research the emphasis was put on very specific symbionts, those that reappear across different generations of the same host and play a very specific role during the host lifetime (normally endosymbionts). The hologenome concept, however, changes the focus of the research and extends it to the whole microbiota. In this vein, the foci of the research are not particular host-symbiont associations, but the association between a host and all of its microbes.
– Second, a fundamental element that frames contemporary discussions about symbiosis is the role of microorganisms for maintaining life as we know it. More concretely, the hologenome concept appeared as a (alleged) conceptual consequence of the observations of the conditions under which healthy corals could grow (i.e. the "coral probiotic hypothesis"). The proponents hypothesized that the best way of explaining health and disease among corals was to propose that corals, with their microbiome, constituted a single unit of selection in evolution. In this sense, the discovery of the physiological relationships between animals and plants and their microorganisms is the basis of the hologenome concept, as well as the basis for understanding its philosophical significance. In fact, this is what distinguishes Schneider's account of the bacterial flora and the account put forwards by defenders of the hologenome concept: while Schneider recognised the existence of an elective affinity between microorganisms and their host, he believed this to be of reduced significance, and thus situated this as a case of "accidental symbiosis". In recent years however, it has been shown that the relationship between a host and its microbiota is not just merely "casual", but that there are very concrete physiological (and developmental) functions that are partially determined and/or realized as a consequence of its presence. This is particularly important because it encourages us to think of the phenomenon's evolutionary possibility. The hologenome concept of evolution is a hypothesis about why this elective affinity, whichis accompanied by the realization of basic functions could have appeared and evolved through time.
– Third, the hologenome concept of evolution, in contrast with previous discoveries made for symbiosis, has caught the attention of many philosophers of biology that had

previously not considered the symbiotic "habit" in much detail. I think there are two reasons for this. The first reason is that, after the important developments in the "omic" sciences, philosophers of biology started paying more attention to microorganisms (O'Malley 2014). In fact, philosophy of biology has been accused of highly ignoring the importance of microorganisms, which despite constituting about 80% of the total biota had not played a significant role in many philosophical disputes (O'Malley and Dupré 2007). This attitude has changed in recent years, and this change is important if we are to understand why the conceptual problems raised by symbiosis research have become more urgent for the philosophers of biology writing at the present. A second important change is the way in which the defenders of the hologenome concept, especially Rosenberg and Zilber-Rosenberg, have framed the debate. In contrast to previous research on symbiosis, which, while acknowledging the evolutionary importance of symbiosis, still treated the phenomenon in ecological terms, the defenders have put emphasis on the evolutionary importance of symbiosis. Particularly, they have provoked one of the most agitated debates among philosophers of biology, the debate about the units of selection. I believe that the emphasis of their understanding of symbiotic assemblages (holobionts) in terms of the units of selection debate has been of special importance for the engagement of philosophers, who have been discussing the issue of units of selection for about half a century. In this vein, framing the debate in terms of units of selection is conceptually novel in relation to previous (nineteenth century) debates.

3 Part III. Concluding remarks

This paper has reviewed some of the current debates about symbiosis and its relation to the problems of biological individuality. It has also traced the historical roots of the current debates, and argued that some of this arise as a consequence of the hologenome concept of evolution, which was were already present to some extent in the nineteenth century, in the context of the original problem of explaining the "living together" of individuals. The review shows how current biological disputes are partially grounded in different philosophical assumptions, but concretely grounded in different conceptions about biological individuality. I have argued that defenders of the hologenome concept tend to emphasize the collaborative aspect of life, and that they show a tendency to focus their studies on the evolution of the symbiotic relationship, irrespective of the different organisms that engage in the symbiosis. On the contrary, detractors of the hologenome concept tend to emphasize the conflicting interests of the entities that compose the holobiont, and, on these grounds, tend to reject any attribution of individuality, conceiving the holobiont as a community of relatively independent individuals. The disagreement among both parties in the dispute is based upon diverging conceptions about biological individuality, as well as upon diverging conceptions about the focal unit of analysis.

Finally, the review has also revealed the connection between the original debates about symbiosis and contemporary debates. I have drawn four parallelisms between the historical and contemporary debates, and emphasized three distinctive issues of the current debates. I have shown how the disagreements amongst both the defenders and the detractors of the holobiontare similar to some of the disagreements of both the defenders and the detractors of symbiosis during the concept's modern inception. In general, the review has shown the existence of an intimate connection between biology, history and philosophy, and how different philosophical assumptions might underlie current debates in biology. Furthermore, I have suggested the importance of the relationship between philosophy and current biological thought, especially concerning the debates on biological individuality, the holobiont and the units of selection, and I have emphasised the historical origin of these debates. I suspect that many current debates in biology are also affected by diverging philosophical assumptions, which have their specific historical background also. Studying these assumptions, as well as their historical sources, is an important and constructive task facilitating firther clarity and understanding on some of these contemporary debates. In this sense, biology, philosophy of biology and history of biology, far from being completely separate disciplines, are totally entangled with one another.

Acknowledgments I would like to thank Staffan Müller-Wille, Sabina Leonelli, Caglar Karaca, John Dupré, José Díez, who read previous versions of this manuscript and made helpful comment. Benjamin Smart is especially acknowledged for all his help and detailed comments in the final version of the manuscript. Finally, I would like to thank two anonymous reviewers for their comments, which clearly helped in improving the content and structure of the paper. This work was economically supported by the Spanish Ministry of Education (FFU16/02570) and the Spanish Ministry of Economy and Competitiveness (FFI2016-76799-P).

References

Booth A (2014) Symbiosis, selection and individuality. Bio Philos 29: 657–673

Bordenstein SR, Theis KR (2015) Host biology in the light of the microbiome: ten principles of holobionts and hologenomes. PLoS Biol. https://doi.org/10.1371/journal.pbio.1002226

Bouchard F, Huneman P (2013) From groups to individuals. Evolution and emerging individuality. The MIT Press, London

Boucher P (1965) Endosymbiosis of animals with plants microorganisms. Interscience Publishers, New York

Boucias DG, Kariithi HM, Bourtzis K, Schneider DI, Kelley K, Miller WJ, Parker AG, Abd-Alla AMM (2013) Transgenerational transmission of the *Glossina pallidipes* Hytrosavirus depends on the presence of a functional Symbiome. PLoS One 8(4):e61150

Brandt K (1881) Über das Zusammenleben von Algen und Tieren. Biologisches Centallblatt 1:524–527

Brucker RM, Bordenstein SR (2012) Speciation by Symbiosis. Trends Ecol Evol 27(8):443–451

Brucker RM, Bordenstein SR (2013) The capacious hologenome. Zoology 116:260–261

Carrapiço F (2015) Can we understand evolution without Symbiogenesis? In: Gontier N (ed) Reticulate evolution: Symbiogenesis, lateral gene transfer, hybridization and infectious heredity. Springer, London, pp 81–106

Chiu L, Eberl G (2016) Microorganisms as scaffolds of biological individuality: an eco-immunity account of the holobiont. Biol Philos 31: 819–837

Clarke E (2010) The problem of biological individuality. Biological Theory 5(4):312–325

Crombie JM (1886) On the algo-lichen hypothesis. Journal of Linnaean Society 21:259–282

Cullen TW, Schofield WB, Barry NA, Putnam EE, Rundell EA, Trent MS, Degnan PH, Booth CJ, Yu H, Goodman AL (2015) Gut microbiota. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. Science 347(6218): 170–175

Dawkins R (1976) The Selfish Gene. Oxford, Oxford University Press

De Bary A (1879) Die Erscheinung der Symbiose. Verlag von Karl J, Trübner

Díaz, JS (2015) El mecanismo evolutivo de Margulis y los niveles de selección. Contrastes: Revista internacional de filosofia 20(1):7–24

DiFrisco J (2017) Kinds of biological individuals: Sortals, projectability, and selection. Br J Philos Sci

Doolittle WF (2017) Darwinizing Gaia. J Theor Biol 434:11–19

Doolittle WF, Booth A (2017) It's the song not the singer: an exploration of holobiosis and evolutionary theory. Biol Philos 32:5–24. https://doi.org/10.1007/s10539-016-9542-2

Douglas AE (2010) The symbiotic habit. Princeton University Press, Oxford

Douglas AE, Werren JH (2016) Holes in the hologenome: why host-microbe symbioses are not holobionts. MBio 7(2):e02099–e02015

Dupré J (2010) The polygenomic organism. Sociol Rev 58(s1):19–30

Dupré J (2012) Processes of life: essays in the philosophy of biology. Oxford University Press, Oxford

Dupré J, O'Malley MA (2009) Varieties of living things: life at the intersection of lineage and metabolism. Philosophy & Theory in Biology 1(December). https://doi.org/10.3998/ptb.6959004.0001.003

Egerton FN (2015) History of ecological sciences, part 52: Symbiosis studies. Bulletin of Ecological Society of America 96(1):80–139

Frank R (1877) Über die biologischen Verthältnisse des Thallus eineger Krustenflecten. Beitrage zur Biologie der Pflanzen 2:123–200

Frank R (1885) Über die auf Wurzelsymbiose beruhende Ernährung gewisser Bäume durch unterirdische Pilze. Berichte der Deutschen Botanischen Gesellschaf 3:128–145

Frank R (2005) On the nutritional dependence of certain trees on root symbiosis with belowground fungi (an English translation of a.B. Frank's classic paper of 1885). Mycorrhiza 15:267–275

Geddes P (1882) Further researchers on animals containing chlorophyll. Nature 25:303–304

Gilbert SF, Epel D (2009) Ecological Developmental Biology. Sinauer Associates

Gilbert SF, Sapp J, Tauber AI (2012) A symbiotic view of life: we have never been individuals. Q Rev Biol 87(4):325–341

Gilbert SF, Rosenberg E, Zilber-Rosenberg I (2017) The holobiont with its hologenome is a level of selection in evolution. In: Gissis SB, Lamm E, Shavit A (eds) Landscapes of collectivity in the life sciences. The MIT Press, London, pp 305–324

Gissis SB, Lamm E, Shavit A (eds) (2017) Landscapes of collectivity in the life sciences. The MIT Press, Cambridge

Godfrey-Smith P (2009) Darwinian populations and natural selection. Oxford University Press, Oxford

Godfrey-Smith P (2015) Reproduction, symbiosis, and the eukaryotic cell. PNAS 112(33):10120–10125

Gontier N (2015) Reticulate evolution: Symbiogenesis, lateral gene transfer, hybridization and infectious heredity. Springer, London

Gontier N (2016a) Symbiosis. In: Kliman RM (ed) The Encyclopaedia of evolutionary biology, vol 4. Academic Press, Oxford, pp 272–281

Gontier N (2016b) Symbiogenesis. In: Kliman RM (ed) The Encyclopaedia of evolutionary biology, vol 4. Academic Press, Oxford, pp 261–271

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O et al (2014) Human genetics shapes the gut microbiome. Cell 159:789–799

Goodrich JK, Davenport ER, Beaumont M, Clark AG, Ley RE (2017) The relationship between the human genome and microbiome comes into view. Annu Rev Genet 51:413–433

Griesemer J (2016) Reproduction in complex life cycles: a developmental reaction norms perspective. Philos Sci 83:803–815

Griesemer J (2017) Landscapes of developmental collectivity. In: Gissis SB, Lamm E, Shavit A (eds) Landscapes of collectivity in the life sciences. The MIT Press, London, pp 25–48

Guerrero R, Margulis L, Berlanga M (2013) Symbiogenesis: the holobiont as a unit of evolution. Int Microbiol 16:133–143

Hester ER, Barott KL, Nulton J, Vermeij MJA, Rohwer FL (2016) Stable and sporadic symbiotic communities of coral and algal holobionts. The ISME Journal 10:1157–1169

Honegger R (2000) Simon Schwendener (1829–1919) and the dual hypothesis of lichens. Bryologist 103(2):307–313. https://doi.org/10.1639/0007-2745(2000)103[0307:SSATDH]2.0.CO;2

Hull DL (1980) Individuality and selection. Annu Rev Ecol Syst 11:311–332. https://doi.org/10.1146/annurev.es.11.110180.001523

Hurst GDD (2017) Extended genomes: symbiosis and evolution. Interface Focus 7:20170001. https://doi.org/10.1098/rsfs.2017.0001

Huttenhower C, Gevers D, Knight R, Creas HH et al (2012) Structure, function and diversity of the healthy human microbiome. Nature 486:207–214

Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ (2010) Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. Science 329:212–215

Khakhina LN (1992) Concepts of Symbiogenesis: a historical and critical study of the research of Russian botanists. Yale University Press, New Haven

Kozo-Polyanski M (1924 [2010]) Symbiogenesis. A new principle in evolution. Edited by V Fett & L Margulis. Cambridge, Harvard University Press

Kropoptkin P (1902) Mutual aid. A factor of evolution. William Heinemann, London

Lamm E (2017) Cultural group selection and Holobiont evolution: a comparison of structures of evolution. In: Gissis SB, Lamm E, Shavit A (eds) Landscapes of collectivity in the life sciences. The MIT Press, London, pp 369–384

Lemanceau P, Blouin M, Muller D, Moënne-Loccoz Y (2017) Let the core microbiota be functional. TRENDS in Plant Science 22 (7): 583–595

Lewontin RC (1970) The units of selection. Annu Rev Ecol Syst 1:1–18

Lidgard S, Nyhart LK (2017) The work of biological individuality. Concepts and contexts. In: Lidgard S, Nyharts LK (eds) Biological individuality. Integrating scientific, philosophical and historical perspectives. The University of Chicago Press, London, pp 17–62

Lipnicki LL (2015) The role of symbiosis in the transmission of some eukaryotes from aquatic to terrestrial environments. Symbiosis 65: 39–53

Lloyd E (2017a) Units and Levels of selection. In EN Zalta (ed.) *Stanford Encyclopaedia of Philosophy*. https://plato.stanford.edu/entries/selection-units/

Lloyd E (2017b) Holobionts as units of selection: Holobionts as interactors, reproducers, and manifestors of adaptation. In: Gissis SB, Lamm E, Shavit A (eds) Landscapes of collectivity in the life sciences. The MIT Press, London, pp 351–367

Lloyd E (2017c) A glimpse of philosophy of biology and collectivities today. In: Gissis SB, Lamm E, Shavit A (eds) Landscapes of collectivity in the life sciences. The MIT Press, London, pp 291–301

Margulis L (1970) The origin of eukaryotic cells. Yale University Press

Margulis L (1990) Words as battle cries – symbiogenesis and the new field of endocytobiology. Bio Sci 40(9):673–677

Margulis L (1991) Symbiogenesis and symbioticism. In: Margulis L, Fester R (eds) Symbiosis as a source of evolutionary innovation. The MIT Press, Cambridge, pp 1–14

Margulis L (1993) Symbiosis in cell evolution: microbial communities in the Archean and Proterozoic eons. WH Freeman and Co., New York

Margulis L (1998) Symbiotic planet. A new look at evolution. Basic Books, New York

Margulis L (2010) Symbiogenesis. A new principle in evolution. Paleontol J 44(12):1525–1539

Margulis L, Fester R (eds) (1991) Symbiosis as a source of evolutionary innovation. The MIT Press, Cambridge

Margulis L, Sagan D (2002) Acquiring genomes. A theory of the origin of species. Basic Books, New York

Martin W, Kowallik K (1999) Annotated English translation of Mereschkowsky's 1905 paper "Über Natur und Ursprung der Chromatophoren im Pflanzanreiche". Eur J Phycol 34(3):287–295

Martin BD, Schwab E (2012) Symbiosis: "living together" in chaos. Studies in the History of Biology 4(4):7–25

Martin BD, Schwab E (2013) Current usage of symbiosis and associated terminology. International Journal of Biology 5:32–45

Maynard-Smith J (1987) Evolutionary progress and levels of selection. In: Dupré J (ed) The latest on the best: essays on evolution and optimality. MIT Press, Cambridge, pp 119–131

Maynard-Smith J (1991) A Darwinian view of symbiosis. In: Margulis L, Fester R (eds) Symbiosis as a source of evolutionary innovation. The MIT Press, Cambridge, pp 26–39

McFall-Ngai M (2015) Giving microbestheir due – animal life in amicrobially dominant world. J Exp Biol 218:1968–1973

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Loso T, Douglas AE, Dubilier N, Eberl G et al (2013) Animals in the bacterial world, a new imperative for the life sciences. PNAS 110(9): 3229–3236

Merezhkowsky C (1905) Über Natur und Ursprung der Chromatophoren imPflanzanreiche. Biologisches Centralblatt 25:593–604

Merezhkowsky C (1910) Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. Biologisches Centralblatt 30:278–303

Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV et al (2016) Cospeciation of gut microbiota with hominids. Science 353:380–382

Moran N, Sloan DB (2015) The Hologenome concept: helpful or hollow? PLoSBiol 13(12):e1002311

O'Malley MA (2017) From endosymbiosis to holobionts: evaluating a conceptual legacy. J Theor Biol 434:34–41. https://doi.org/10.1016/j.jtbi.2017.03.008

O'Malley MA, Dupré J (2007) Size doesn't matter: towards a more inclusive philosophy of biology. Biol Philos 22:155–191

Ochman H, Worobey M, Kuo C-H, Ndjango N-BN, Peeters M et al (2010) Evolutionary relationships of wild hominids recapitulated by gut microbial communities. PLoS Biol 8(11):e10000546

Okasha S (2006) Evolution and the levels of selection. Oxford University Press, Oxford

Oldroyd GED (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nat Rev Microbiol 11:252–263

O'Malley MA (2014) Philosophy of microbiology. Cambridge University Press, Cambridge

Oulhen N, Schulz BJ, Carrier TJ (2016) English translation of Heinrich Anton de Bary's 1878 speech, 'die Erscheinung der Symbiose'('De la symbiose'). Symbiosis 69:131–139. https://doi.org/10.1007/s13199-016-0409-8

Paracer S, Ahmadjian V (2000) Symbiosis: an introduction to biological associations. Oxford University Press, Oxford

Peacock KA (2011) Symbiosis in ecology and evolution. In: Gabbay DM, Thagard P, Woods J (eds) Handbook of the philosophy of science: philosophy of ecology. North Holland, San Diego, pp 219–250

Portier P (1918) Les Symbiotes. Masson, Paris

Pound R (1893) Symbiosis and mutualism. Am Nat 27(318):509–520

Pradeu T (2016a) The many faces of biological individuality. Biol Philos 31:761–773

Pradeu T (2016b) Organisms or biological individuals? Combining physiological and evolutionary individuality. Biol Philos 31:797–817

Queller DC, Strassmann JE (2009) Beyond society: the evolution of organismality. Philos Trans R Soc B 364:3143–3155

Queller DC, Strassmann JE (2016) Problems of multispecies organisms: endosymbionts to holobionts. Biol Philos 31:855–873

Relman DA (2012) Microbiology: learning about who we are. Nature 486:194–195

Reshef L, Koren O, Loya Y, Zilber-Rosenberg I, Rosenberg E (2006) The coral probiotic hypothesis. Environ Microbiol 8:2068–2073

Rosas-Pérez T, Vera-Ponce de León A, Ramírez-Puebla ST, Rincón-Rosales R, Martínez-Romer J, Dunn MF, Kondorosi E & Martínez-Romero E (2017) The Symbiome of Llaveia Cochineals (Hemiptera: Coccoidea: Monophlebidae) Includes a Gammaproteobacterial Cosymbiont Sodalis TME1 and the Known Candidatus Walczuchella monophlebidarum. In VDC Shields (ed.): *Insect Physiology and Ecology*. DOI: https://doi.org/10.5772/66442. Available from: https://mts.intechopen.com/books/insect-physiology-and-ecology/the-symbiome-of-llaveia-cochineals-hemiptera-coccoidea-monophlebidae-includes-a-gammaproteobacterial

Rosenberg E, Zilber-Rosenberg I (2014) TheHologenome concept. Springer, London

Rosenberg E, Zilber-Rosenberg I (2016) Microbes drive evolution of animals and plants: the hologenome concept. MBio 7(2):e01395–e01315

Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I (2007) The role of microorganisms in coral health, disease and evolution. Nat Rev Microbiol 5:355–362

Rosenberg E, Sharon G, Atad I, Zilber-Rosenberg I (2010) The evolution of animals and plants via symbiosis with microorganisms. Environ Microbiol Rep 2(4):500–506

Roughgarden J, Gilbert SF, Rosenberg E, Zilber-Rosenberg I & Lloyd EA (2017). Holobionts as units of selection and a model of their population dynamics and evolution. *Biological Theory*

Sagan L (1967) On the origin of mitosing cells. Journal of Theoretical Biology 14: 225–274

Sapp J (1994) Evolution by association. A history of symbiosis. Oxford University Press, New York

Sapp J (2002) Paul Buchner (1886-1978) and hereditary symbiosis in insects. Int Microbiol 5(3):145–150

Sapp J (2003) Genesis: the evolution of biology. Oxford University Press, New York

Sapp J (2004) The dynamics of symbiosis: an historical overview. Can J Bot 82:1046–1056

Sapp J (2010) Saltational symbiosis. Theory Biosciences 129:125–133

Sapp J, Carrapiço F, Zolotonosov M (2002) Symbiogenesis. The hidden face of Constantin Merezhkowky. History and Philosophy of the Life Sciences 24(3/4):413–440

Schneider A (1897) The phenomena of Symbiosis. Minnesota Botanical Studies 1(9):923–948

Schwendener S (1868) Über die Beziehungen zwischen Algen und Flechtengonidien. *Botanische Zeitung* [Berlin]: 289–292

Shropshire JD, Bordenstein SR (2016) Speciation by symbiosis: the microbiome and behavior. MBio 7(2):e01785–e01715

Skillings D (2016) Holobionts and the ecology of organisms: multi-species communities or integrated individuals? Bio Philos 31:875–892

Sommer F, Bäckhed F (2013) The gut microbiota – masters manipulator of host development and physiology. Nat Rev Microbiol 11(4):227–238

Spencer H (1899) The principles of biology. D. Appleton & Co., New York

Stahl E (1877) *Beiträge zur Entwickelungsgeschichte der Flechten* (vols. 1 & 2). Leipzig: A Felix

Stencel A (2016) The relativity of Darwinian populations and the ecology of endosymbiosis. BiolPhilos 31:619–637

Taxis TM, Wolff S, Gregg SJ, Minton NO, Zhang C, Dai J, Schnabel RD, Taylor JF, Kerley MS, Pires JC, Lamberson WR, Conant GC (2015) The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. Nucleic Acids Res 43(20):9600–9612

Theis KR, Dheilly NM, Klassen JL, Brucker RM, Baines JF, Bosch TCG, Cryan JF, Gilbert SF, Goodnight CJ, Lloyd EA, Sapp J, Vandenkoornhuyse P, Zilber-Rosenberg I, Rosenberg E, Bordenstein SR (2016) Getting the hologenome concept right: an eco-evolutionary framework for hosts and their microbiomes. mSystems 1(2):e00028–e00016

Trappe JM (2005) A. B. Frank and mycorrhizae: the challenge to evolutionary and ecologic theory. Mycorrhiza 15(4):277–281

Tripp EA, Zhans N, Schneider H, Huang Y, Mueller GM, Hu Z, Häggblom M, Bhattacharya D (2017) Reshaping Darwin's tree: impact of the symbiome. TRENDS in Ecology and Evolution 32(8): 552–555

Turpin W, Espín-García O, Xu W, Silverberg MS, Kevans D, Smith MI, Guttman DS, Griffiths A et al (2016) Association of host genome with intestinal microbial composition in a large healthy cohort. Nat Genet 48(11):1413–1417

Van Beneden P-J (1876) Animal parasites and messmates. Henry S. King, London

Wallin IE (1927) Symbioticism and the origin of species. Williams & Wilkins Co., Baltimore

Wilkinson DM (2001) At cross purposes. Nature 412:485

Wilson RA, Barker M (2013) The biological notion of individual. In EN Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2017/entries/biology-individual/

Zilber-Rosenberg I, Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenometheoryof evolution. FEMS Microbiol Rev:723–735

Zook D (2015) Symbiosis: Evolution's co-author. In: Gontier N (ed) Reticulate Evolution. Springer, London, pp 41–80

teorema
Vol. XXXVIII/3, 2019, pp. 95-120
ISSN: 0210-1602
[BIBLID 0210-1602 (2019) 38:3; pp. 95-120]

Equilibrium Explanation as Structural Non-Mechanistic Explanations: The Case of Long-Term Bacterial Persistence in Human Hosts

Javier Suárez and Roger Deulofeu

RESUMEN

Philippe Huneman ha cuestionado recientemente los límites en la aplicación de los modelos mecanicistas de la explicación científica en base a la existencia de lo que denomina "explicaciones estructurales", en las que el fenómeno se explica en virtud de las propiedades matemáticas del sistema en que el fenómeno ocurre. Las explicaciones estructurales pueden darse en formas muy diversas: en virtud de la forma de *pajarita* (*bowtie*) de la estructura, de las propiedades topológicas del sistema, de los equilibrios alcanzados, etc. El papel que juegan las matemáticas en las explicaciones que apelan a la estructura de pajarita o a las propiedades topológicas del sistema ha sido recientemente examinado en varios trabajos. Sin embargo, el papel exacto que juegan las matemáticas en el caso de las explicaciones en términos de equilibrio aún no ha sido totalmente clarificado, y diferentes autores defienden interpretaciones contradictorias, algunas de las cuales las asemejarían más al modelo defendido por algunos filósofos mecanicistas que al modelo estructural de Huneman. En este trabajo, tratamos de cubrir ese déficit estudiando el papel que juegan las matemáticas en el modelo de equilibrio anidado (*nested equilibrium*) elaborado por Blaser y Kirchner para explicar la estabilidad de las asociaciones ontogenética y filogenéticamente persistentes entre humanos y microorganismos. De nuestro análisis se desprende que su modelo es explicativo porque i) se identifica una estructura matemática del sistema que viene dada por un conjunto de ecuaciones diferenciales que satisfacen una estrategia evolutivamente estable; ii) la estructura anidada del modelo hace que la estrategia evolutivamente estable sea robusta ante posibles perturbaciones; iii) esto es así porque las propiedades del sistema empírico son isomorfas a, pero no causalmente responsables de, las propiedades de la estrategia evolutivamente estable. La combinación de estas tres tesis hace que las explicaciones en términos de equilibrios se asemejen más al modelo estructural de explicación que al modelo mecanístico.

PALABRAS CLAVE: *explicación científica; mecanismos; explicación en términos de equilibrio; explicaciones estructurales; explicaciones no causales; estrategia evolutivamente estable*.

ABSTRACT

Philippe Huneman has recently questioned the widespread application of mechanistic models of scientific explanation based on the existence of structural explanations, i.e. explanations that account for the phenomenon to be explained in virtue of the mathematical properties of the system where the phenomenon obtains, rather than in terms of the mechanisms that causally produce the phenomenon. Structural explanations are very di-

verse, including cases like explanations in terms of bowtie structures, in terms of the topological properties of the system, or in terms of equilibrium. The role of mathematics in bowtie structured systems and in topologically constrained systems has recently been examined in different papers. However, the specific role that mathematical properties play in equilibrium explanations requires further examination, as different authors defend different interpretations, some of them closer to the new-mechanistic approach than to the structural model advocated by Huneman. In this paper, we cover this gap by investigating the explanatory role that mathematics play in Blaser and Kirschner's nested equilibrium model of the stability of persistent long-term human-microbe associations. We argue that their model is explanatory because: i) it provides a mathematical structure in the form of a set of differential equations that together satisfy an ESS; ii) that the nested nature of the ESSs makes the explanation of host-microbe persistent associations robust to any perturbation; iii) that this is so because the properties of the ESS directly mirror the properties of the biological system in a non-causal way. The combination of these three theses make equilibrium explanations look more similar to structural explanations than to causal-mechanistic explanation.

KEYWORDS: *Scientific Explanation; Mechanisms; Equilibrium Explanations; Structural Explanations; Non-Causal Explanations; Evolutionarily Stable strategy*.

In the last few years, a new trend in the debates about scientific explanation has flourished in philosophy of science. This new trend, "new-mechanism," emphasizes the role of mechanisms in scientific discourse in general, and in scientific explanation in particular [Machamer et al. (2000); Glennan & Illari (2017)]. Inspired by the developments in molecular biology, new-mechanists redefine causalism and argue that to explain a phenomenon consists in providing the mechanism that produces it. In the new-mechanist tradition, mechanisms are taken to be a set of *entities* (parts) and *activities* (operations) with a particular *organization* such that their causal interactions bring the phenomenon to be explained about [Glennan (2002); Bechtel & Abrahamsen (2005); Craver & Darden (2013); Craver (2007); Nicholson (2012); Issad & Malaterre (2015); Deulofeu & Suárez (2018)]. Thus, for a scientific explanation to be mechanistic, it must fulfill two necessary and sufficient conditions. First, it must identify a *model of mechanism* in which the mechanism is individuated by its parts, operations and organization. Second, it must provide a story of how the components of the mechanism are causally connected in such a way that they produce the *explanandum*.

New-mechanists share a basic commitment to a causal view of the world combined with: 1) the rejection of the Hempelian idea that explanations take the form of logical arguments, either inductive or deductive, and 2) the notion that mechanisms provide the causal "ingredient" that scientific explanations require to be genuinely explanatory[1]. Furthermore, they often assume a hierarchical view of mechanisms, acknowledging the existence of a diversity of scientific explanations in every science, thus

neither renouncing to the explanatory role of the special sciences, nor to the possible existence of mechanistic inter-level (hierarchical) explanations among different sciences [Krickel (2018)].

The wide scope of the New Mechanism account of scientific explanation in biology has been questioned due to the existence of explanations that seem to lack the causal ingredient that new-mechanists demand. One of the traditional explanatory types where this happens is in equilibrium explanations, where the mathematical properties of the empirical system (i.e. the fact that it reaches an equilibrium point) are taken as explanatory, irrespectively of the causal-mechanistic details of the system. Starting with Sober (1983), equilibrium explanations have been hypothesized to constitute an alternative to purely causal-mechanistic explanations [Batterman & Rice (2014); Rice (2015); Huneman (2018b), (2018c)]. However, it has also been argued that some equilibrium explanations admit a causal interpretation, if "causality" is understood in Woodward's interventionists terms [Woodward (2003); Kuorikoski (2007); Potochnik (2015)]. If the later were the case, as some new-mechanists are committed to an interventionist Woodwardian view of causation [Craver (2007); Kaplan & Craver (2011)], it could be argued: first, that the mathematical components that are present in equilibrium explanations describe the causal relationships among the entities of the system; second, that equilibrium explanations do not then constitute a real exception to the new-mechanist trend. The existence of these contradictory interpretations of the nature of equilibrium explanations (causal vs. non-causal) creates an important gap to understand how they gain their explanatory force, as well as about the specific role of causality in scientific explanation: is causality — at some level — a necessary ingredient in every scientific explanation, or are non-causal explanations also legitimate in certain cases?

In this paper, we aim to clarify this issue by studying Blaser & Kirschner's (2007) nested equilibrium model (NEM, hereafter) of the persistence of bacteria in human hosts. Our choice of this case is motivated by two reasons: on the one hand, Blaser & Kirschner's NEM explains the phenomenon in terms of the existence of an evolutionarily stable strategy (ESS, hereafter) among the different interacting organisms, a feature that makes it sufficiently analogous to most cases of equilibrium explanations reviewed in the philosophical literature so that our conclusion can shed light on the nature of scientific explanation; on the other hand, the explanatory force of their model is also conditional on the existence of a nestedness among different biological scales, i.e. on the

existence of a hierarchy of interrelated ESSs. As the acknowledgment of the existence of a hierarchy of mechanisms is a hallmark of the new-mechanist account of scientific explanation, and, to our knowledge, cases of nested equilibria have never been studied before in the philosophical literature, we believe that our case study could bring new light to the study of the old phenomenon of equilibrium explanations. Our aim is thus to analyse the explanatory role that the appeal to the existence of equilibria at different levels plays in the NEM. In that vein, we intend to provide a better understanding of the nature of equilibrium explanation, and to the role of causality in scientific explanation[2]. To do so, we frame the paper in the context of the debate between Huneman's structural account of scientific explanation and the causal-mechanistic account.

In section I, we introduce the general account of structural explanations presented by Huneman (2018a) and motivate the necessity of discussing the precise nature of equilibrium explanation to understand whether, and if so, to what extent, equilibrium explanations fit Huneman's account, or are rather a special case of causal-mechanistic explanations. In section II, we present our case study. In section III, we present our philosophical analysis. We first argue that the explanatory force of Blaser & Kirschner's NEM is mainly provided by the concept of ESS, plus the mathematical modelling that defines each strategy at each of the levels of the hierarchy, rather than by the causal-mechanistic details of the system. Additionally, the nested nature of the different ESSs plays a role in making the system robust to every possible intervention at different levels. Thirdly, and connected to this last point, we argue that no role is left for any causal element in their model, thus suggesting that their explanation constitutes a case of structural explanation as Huneman has defined it. Finally, in section IV, we present our conclusions.

I. Explaining With and Without Causes: The Role of Mathematics in Equilibrium Explanations

In recent years, the universal application of the "new-mechanist" account of scientific explanation in biology has been questioned on the basis of the existence of a family of explanations that do not rely on any causal features of the system whose properties they explain, but rather on its mathematical properties [Huneman (2010), (2018a), (2018b); Woodward (2013); Rice (2015); Kostic (2018), (2019); Deulofeu et al. (2019)]. Huneman has called these explanations "structural", and defines them as follows:

> Family of explanations for which the mathematical tools used in the description of an explanandum system belong to a mathematical structure whose properties are directly explanatory of some aspects of the system (such as equilibria, behaviour, limit regime, asymptotic behaviour, etc.) (…) They explain by accounting for the explananda through pinpointing structural relations that are mathematical relations of some sort. Mathematics here are not representing a dependence between structures in the world, but they are constituting the structural dependence itself, (…) and in virtue of that they are explanatory [Huneman (2018a), p. 695].

In contrast with mechanistic explanations, structural explanations do not include any mechanism, nor any causal story in their *explanans*. Furthermore, the inclusion of any of these elements would usually be taken as counterproductive to account for the *explanandum*. Structural explanations are abundant in systems biology, where an extensive amount of data has to be interpreted by using mathematical and computational tools [Green (2016), (2017); Green & Jones (2017); Brigandt et al. (2017)]. Huneman explicitly argues that some of the properties of the biological systems studied under the label of "systems biology" can only be explained by appealing to the formal (mathematical) properties that characterize those systems. A well-known example of this, studied by Jones (2014), is the vulnerability of the immunological system to attacks to the CD4+ T-cells. Drawing upon Kitano & Oda's (2006) case study, Jones argues that what explains the vulnerability of the human immune system to attacks on this particular component is its bowtie structure: because the human's immune system has a bowtie structure such that CD4+ T-cells are non-redundant elements in the core of the bowtie, the system is vulnerable to attacks on this type of cells (Figure 1). What is more important is that the vulnerability to attacks on CD4+ T-cells is not a consequence of the causal-mechanistic processes that produce the vulnerability: it is a consequence of the topological properties of the architecture (organization) of the immunological system. These topological properties determine its vulnerability to attacks on its core, as it is the only non-redundant element of the network, which is furthermore a necessary step for every other immunological process. Huneman summarized this kind of explanation as follows: "what is epistemically proper to this network modelling is that the topological properties found in the networks are such that they explain some of the properties one is interested in [vulnerability to attacks on CD4+ T-cells], (…) the instantiation of these properties is explained by the fact that the network is of such topological nature" [Huneman (2018b) p 127].

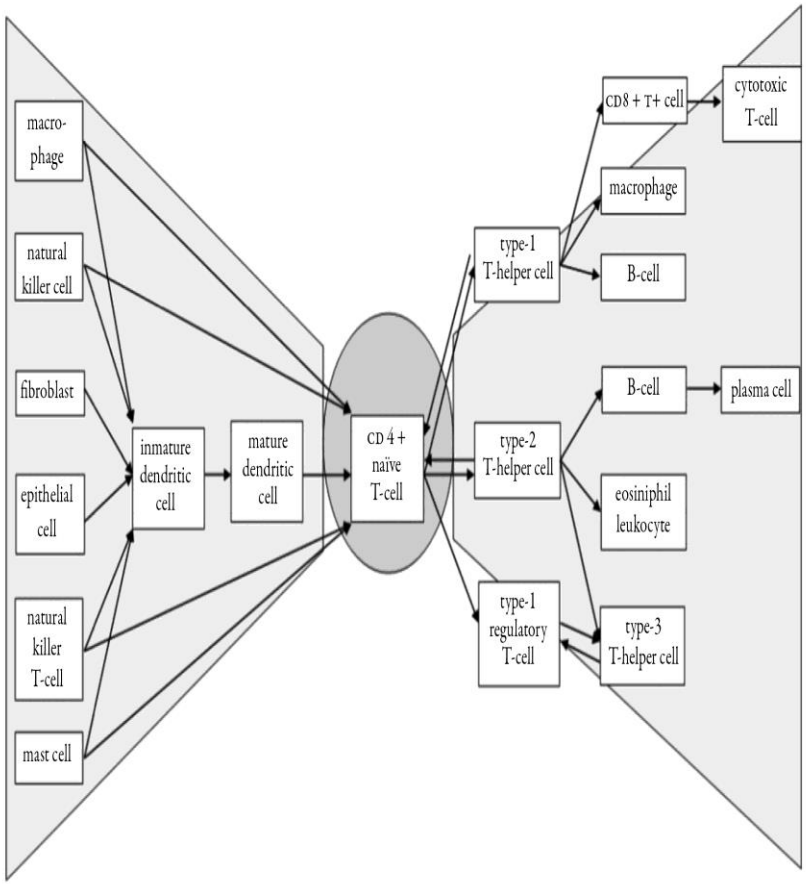

FIGURE 1. Bowtie structure of the immune system, with the CD4+ T-cells in the core of the bowtie. From Jones (2014), p. 1138, Fig. 1.

A second point that is epistemically proper to this kind of explanation is that the mechanisms that "sustain" the realization of such topological properties are irrelevant for explaining those properties (namely, the vulnerability of the network) [Huneman (2018c) pp. 6-8; Deulofeu et al. (2019); Moreno & Suárez, (submitted)][3].

Structural explanations are not restricted to cases of topological explanation, though. In his (2018c), p. 6, Huneman outlines the case of explanations in microeconomics, particularly the "ice cream vendors" problem — a direct application of the theory of Nash equilibrium to

human behaviour. In this situation, we imagine that there are two vendors standing on a beach and need to decide where to situate their stall in order to maximize their sales. Microeconomics, relying on game theory, says that the vendors will situate their stall in the middle of the beach, next to each other, to attract customers both in the area around them and in their extremes. By placing themselves in the middle of the beach, the vendors generate a Nash equilibrium, a situation where none of the players (the vendors) can change their strategy without decreasing their benefits (potential customers). Let us suppose we have to explain a scenario where there are two vendors placed in the middle of the beach. What explains the fact that both of them place their stalls in the middle? Huneman replies: "the fact that it simultaneously maximizes the share of each of them, or in other words, that it instantiates a Nash equilibrium." And adds: "[t]he mechanisms through which vendors move, decide, sell or buy, etc. are not explanatory relevant to this precise question" [Huneman (2018c), p. 6].

Nonetheless, Huneman just sketches the elements that make the Nash equilibrium explanatory in the case of the "ice cream vendors" but does not specify in detail what explaining with equilibria exactly entails, nor what is his reason to believe that mechanisms do not play any explanatory role in equilibrium explanations. Previous analyses of the role of equilibria in scientific explanations had been presented in Sober (1983) and Kuorikoski (2007). However, both authors reach opposing conclusions about where equilibrium explanations gain their explanatory force from: while the former argues that "equilibrium explanations show how the cause of an event can be (statistically) *ir*relevant to its explanation", and that their explanatory force comes exclusively from their mathematical structure [Sober (1983), p. 201], the latter believes that "explanations of singular events are indeed causal, even those supplied by equilibrium models" [Kuorikoski (2007), p. 149]. These opposing conclusions are interesting because they leave open whether equilibrium explanations must be considered a subtype of structural explanation (Sober), or a subtype of causal-mechanistic explanation (Kuorikoski), thus creating an important gap in how to understand the role of mathematics in this type of explanation. In addition to that, they leave open a question about the role of causality in scientific explanation in general for, if as Kuorikoski argues, even equilibrium explanations are in the end causal, then it could be argued that causality is a necessary ingredient in every genuine case of scientific explanation.

In the next section, we introduce Blaser & Kirschner's NEM of the persistence of bacteria in human hosts as a case study that we will use to motivate our response to these two questions.

II. A Nested Equilibrium Explanation of the Persistence of Bacteria in Human Hosts

Humans harbour an abundant number of microbes in their guts that constitute the human microbiome [Huttenhower et al. (2012); Lozupone et al. (2012)][4]. Among those microorganisms, some persist in our guts throughout our entire whole life cycle, whereas others are mainly transient, or appear in specific moments of our development, disappearing afterwards [Chiu & Gilbert (2015)]. Furthermore, some of those are hypothesized to have established long-term associations with humans over millions of years, with some people speculating that they might constitute co-evolved systems or hologenomes [Rosenberg & Zilber-Rosenberg (2014), (2016); Díaz (2015); Suárez (2018); Suárez & Triviño (2019); cf. Moran & Sloan (2015); Douglas & Werren (2016)]. Irrespectively of the evolutionary nature of those associations, the fact that organisms from different species engage in persistent long-term associations with each other is paradoxical from the perspective of the neo-Darwinian model of life and evolution. According to this model, when two individuals of different species associate, i.e. when they share the same habitat or niche, each one will pursue its own fitness interests. In this scenario, it might happen that the two organisms coexist peacefully for a period of time but, normally, peaceful coexistence will tend to break down: on the one hand, in the moment in which an opportunity for one of the organisms to benefit in detriment of the other appears, it will tend to grow to maximize its fitness until the other organism is destroyed (appearance of cheaters); on the other hand, it is also not infrequent that in a stable biological population where one out of two different survival strategies has been adopted among the members, the population becomes invaded by individuals that adopt an alternative strategy, until the point where the population collapses (external invasion). For these reasons, peaceful associations among organisms of different species are rare and will normally be short-term. Then, how is it possible that humans and some of their microbes establish persistent infections that are not disrupted by cheaters[5]? And which are the mechanisms that allow long-term associations that survive the challenges of sharing a habitat and are not perturbed by external invaders?

Blaser and Kirschner have recently developed a model "*to explain* the common features of microbial persistence in their human hosts" [(2007), p. 847, emphasis added)], i.e. to explain why humans and some specific microorganisms have overcome the difficulties of co-habitation[6]. They speculate that those situations represent a successful phenotype that must be maintained according to certain eco-evolutionary rules. In their view:

> persistence represents the evolved selection for balancing host and microbial interests, resulting in an equilibrium that, by definition, is long-term but not necessarily forever stable. We hypothesize that maintenance of this equilibrium requires a series of evolved, nested equilibria to achieve the overall homeostasis [Blaser & Kirschner (2007), p. 843].

They argue that such nested equilibria will be observed at different timescales: microscopic, at the level of the interactions between the immunological system of the host and cell-receptors of the microbes; mesoscopic, at the level of tissue function; tissue in which the microbe population inhabits; macroscopic, where evolutionary changes in the host and the microbe will occur to guarantee microbe transmission[7]. Blaser and Kirschner believe that any of these levels conforms to Nash equilibria in the form of an ESS that allows the persistence of the relationship. This is so because both the host and the microorganism will have developed a very specific hierarchy of cross-signalling mechanisms that generate a set of positive and negative feedback loops with each other that guarantee that the overall equilibrium is not disrupted.

Blaser and Kirschner's model begins by defining five populations at the microlevel whose changes with respect to certain variables are followed over time [see also Blaser & Kirschner (1999); Blaser & Atherton (2004); Blaser (2006)]. In the case of *Helicobacter pylori*, the variables include: *M*, which represents the population of mucus-living *H. pylori* (rate of change); *A*, which represents the *H. pylori* population that adhere to epithelial cells; *N*, which represents the concentration of nutrients available to bacteria derived from inflammation; *E*, which represents the concentration of effector molecules (molecules that the microbes generate to achieve some aims, such as suppressing immune response by the host); and *I*, that stands for the host response. Blaser and Kirschner's NEM includes five differential equations that track the changes in the variables of their model, as well as how they interact with each other[8].

For instance, to study how the concentration of mucus-living *H. pylori* varies over time due to the interaction with the other populations, they introduce the following differential equation:

$$\frac{dM}{dt} = g_m \alpha N(t) - \mu_m M(t) - \alpha M(t)(K - A(t)) + \delta A(t) \qquad (1)$$

where, g_m, α, μ_m and δ are parameters, whose value will depend on the situation; N, M, A (mentioned above) and K (the epithelial carrying capacity) are variables that together will determine the rate of change of the mucus-living population M. In (1), $g_m \alpha N(t) M(t)$ represents the potential growth of the population in virtue of the nutrient availability; $\mu_m M(t)$, represents the loss of *H. pylori* due to the process of mucus shedding; and $\alpha M(t)(K - A(t)) + \delta A(t)$ represents the potential loss/gain of *H. pylori* due to migration between the epithelial and the mucus-living populations. Obviously, migration from M to A can only happen when $A < K$, namely, when there is still room for more adherence to epithelial cells, and the opposite is the case for migration from A to M. Adherent sites are always limited or otherwise *H. pylori* would grow too much, risking the stability of the symbiotic association.

The inflammation induced by the bacteria on the host is captured by measuring the change of nutrient concentration over time:

$$\frac{dN}{dt} = \frac{b}{(b + I(t))} \beta E(t) - g_m N(t) M(t) - g_\alpha N(t) A(t) \qquad (2)$$

In (2), b, β, g_m and g_α are parameters. $N(t)$ is characterized by a gain term that is a function of the concentration of effector molecules, E, and the host response I. The equation shows the direct proportionality that exists between E and N, and the inverse proportionality between I and N. In other words, it shows the limiting effect that the host response has over the nutrient concentration, as well as the inducing effect of the bacteria on the nutrient concentration. (2) also specifies the rate of assimilation of nutrients of the mucus-living bacterial population and of the adherent epithelial populations.

Furthermore, for a microbe-host association to be *evolutionarily* persistent, the microbe needs to develop strategies for transmission. R_o captures this concept, quantifying "the transmission potential of a microparasite as the average number of secondary infections occurring when a single infectious host is introduced into a universally susceptible host population" [Blaser & Kirschner (2007) p. 844].

$$R_o = \frac{BN}{(x + b + v)} \tag{3}$$

In (3), *BN* measures the transmission rate as a function of the population size, *x* measures the rate of host mortality due to the microbe (measure of virulence), *b* is the rate of mortality of the host population independently of the microbe (measure of lifespan), and *v* is the rate at which the host recovers from the microbe infection (measure of immunity). Usually, for $R_o > 1$ microbial transmission is sustained whereas for $R_o < 1$ microbial transmission goes extinct.

Blaser and Kirschner show that in a persistent microbe-host association those five differential equations remain constant, and any deviation in one of the equations gets immediately counter-balanced by the adjustment of the other equations, keeping the equilibrium stable. Thus, Blaser and Kirschner claim this can only be possible if the system behaves according to a Nash equilibrium, and if the strategies followed by microbe and host conform to an ESS. Let us now see how an ESS can account *explanatorily* for observed constancy.

II.1. *The Role of the Evolutionarily Stable Strategy in Blaser and Kirschner's Model*

Nash equilibrium is a very common situation in game theory. It obtains when two players in a non-cooperative game adopt a strategy such that no individual change will render greater benefits to any of them, i.e. such that every change in the strategy that one of the players adopts independently will result in lower individual profit for that player. Nash equilibria are not necessarily, however, optimal strategies. It is sometimes possible to obtain a better net result if both players change their strategy simultaneously and a new equilibrium is reached. Nonetheless, this will only occur if *both partners* modify their strategy co-ordinately, but not if they do so independently. Therefore, no player has any incentive to modify his strategy individually. The prisoner's dilemma constitutes a typical example of a game whose solution is provided by a Nash equilibrium (Table 1). In this situation, two individuals — A and B — are accused independently of a crime, and each of them is interrogated separately and offered a deal: 1) if A betrays B and accuses her of having committed the crime, while B stays silent, A will have 4-years reduction of sentence and B will have no reduction (and the same, but inverted, occurs if B betrays A while A remains silent); 2) if both stay silent, each

of them will have a 3-years reduction of sentence; 3) if both betray each other, each will have a 1 years reduction of sentence. In this scenario, the Nash equilibrium is reached in situation 3), when both players betray each other. Of course, the result that they obtain is not optimal (each of them will only get 1 year reduction of sentence), but is such that none of them has any incentive to change her strategy individually, unless the other also does so, as otherwise she will have a bigger individual cost, i.e. she will have less years of reduced sentence [Nash (1950a), (1950b); Gintis (2000)].

A \ B	Betrays	Remains silent
Betrays	**1, 1**	4, 0
Remains silent	0, 4	**3, 3**

TABLE 1. Payoff matrix for the prisoner's dilemma. The numbers represent the amount of years that each subject would have as reduction of sentence. The optimal strategy is that where both remain silent (italics). Only the strategy where both betray constitutes Nash equilibria (bold).

An ESS is a biological strategy that, when it is adopted in a population, natural selection alone will keep the population safe from "intruder populations", in so far as the organisms that adopt an alternative strategy will be selected against. All ESSs are cases of Nash equilibria, but the opposite is not the case. If a solution to a non-cooperative game represents Nash equilibrium that is not an ESS, the solution could be disrupted by an alternative strategy that drives the population towards an alternative Nash equilibrium that constitutes an ESS [Smith & Price (1973); Smith (1974); Easley & Kleinberg (2010), pp. 209-227]. For instance, take the case of the stag hunt game (Table 2). This is a two players' game, where each player has two possible exclusive strategies: hunt-hares or hunt-stags. In this situation, there are three possible scenarios: 1) that both individuals are hare-hunters (case where both obtain a fitness benefit of 2); 2) that both individuals are stag-hunters (both obtain a fitness benefit of 3); 3) that one of the individuals is a hare-hunter whereas the other is a stag-hunter (in which case the hare-hunter obtain a fitness benefit of 3, whereas the stag-hunter obtains a fitness benefit of

0). In this situation, strategies 1) and 2) constitute a Nash equilibrium, for none of the players could get a better payoff by changing strategy. However, only 1) constitutes an ESS: while a hare-hunter and a stag-hunter do equally well when they are paired with a stag-hunter (fitness benefit of 3), hare-hunters score better than stag-hunters when they are paired with hare-hunters (hare-hunters score 2, while stag-hunters score 0). That means the stag-hunting strategy is not an ESS because if a hare-hunter is introduced in a population of stag-hunters, the population will evolve towards a population of hare-hunters. On the other hand, a population where all the individuals are hare-hunters represents an ESS, because if a stag-hunter is introduced in the population, it will be eventually extinct, for its fitness benefit will be lower than the fitness benefit of hare-hunters.

	Stag-hunter	Hare-hunter
Stag-hunter	**3, 3**	0, 3
Hare-hunter	3, 0	***2, 2***

TABLE 2. Payoff matrix for the stag hunt game. The numbers represent the net benefit for the individuals in the population that engage in the game. Cases where all the individuals in the population hunt exclusively stags or exclusively hares represent Nash equilibria (bold). However, only the case where both individuals hunt hares represent an ESS (italics).

Blaser and Kirschner apply this type of reasoning to persistent long-term host-microbe associations to argue that the situation must be the one that is obtained in Nash equilibrium, particularly in ESSs, where both positive and negative feedback between the host and the microbe occur, so that the equilibrium persists over time. The core idea of their model is that the equilibrium obtained at the microscopic level immediately affects the equilibrium at superior levels (mesoscopic and macroscopic). At the same time, the equilibrium at the higher levels affects in a specific way the possibility of new microbe-host persistent associations. The equilibria are nested and the association does not get in principle disrupted. The interaction among levels, partially captured by the equations (1)-(3), is as follows:

first, on the microscopic level one would find the microbial population, localized on an organ or tissue of the host, and the population of immune host cells responsible of recognizing the microbe population. The structure of both populations will depend on the nature of the original founder strain, the possibility for generating genetic variants, the selective pressures from other microbial cells in the same tissue and, more importantly, from the selection that the persistent microbe and the immune cells exert on each other [e.g. (Pradeu et. al 2013); Pradeu & Vivier (2016); Eberl (2016)]. The nature of the interactions between the organisms in the microscale will shape tissue function (or malfunction), and thus will partially determine the viability of the host, as well as the opportunity for microbial transmission (mesoscale). Finally, the effects of the microbe on the viability of the host will determine the host population structure (macroscale) that in return will affect microbial transmission (mesoscale) (Figure 2).

Even if the model illustrated in Figure 2 looks like a multilevel mechanism, for it appeals to a model of mechanism, it lacks the adequate type of causal stories that new-mechanists demand to have a proper explanation. First, because multilevel causation is mysterious, as Craver and Bechtel illustrate (2007), since causal relations happen exclusively intra-level. Second, because the type of inter-level readjustments of the system are symmetrical, occurring both top-down (e.g. from the macroscale to the mesoscale, or from the latter to the microscale), and bottom-up (e.g. from the microscale to the mesoscale, or from the latter to the macroscale), while relations between cause and effect are always asymmetrical. Third, because even if there could be a way to capture inter- and intra-level causal relations, this would be at odds with the information that NEM conveys and appeals to. NEM does not specify the causal way in which the entities at one level affect the entities at another level. It only specifies that the disruption of the equilibrium at one level will either prompt the collapse of the system (i.e. its death), or it will prompt the re adjustment of the equilibrium at that level due to the equilibria that exist in the other scales. In other words, NEM is not specific about how the equilibrium will be readjusted, it only predicts that it will be readjusted, provided that the other levels keep their equilibrium states. The causal elements (if any) that will bring this readjustment are irrelevant for the explanation of this behaviour in terms of NEM. What matters is exclusively the nested structure of the host-symbiont system (see section 4 for the full details).

In that vein, the nested structure of the model and the level of complex interactions between the different elements at the three scales (Nash equilibria, ESS) grant the persistence of the association. As it was

said before, one of the reasons why host-microbe associations do not normally last long is due to the presence of cheaters, organisms that enjoy the profits of the associations without paying the cost. Nash equilibria avoid the appearance of cheaters: cheaters are players that change their strategy unilaterally; in Nash equilibria, every player that does so is condemned to failure, and thus will be removed from the population. Furthermore, as the Nash equilibria that are reached in the population adopt the form of an ESS, it is not possible that an external invader adopting an alternative strategy disrupts the persistence of the association.

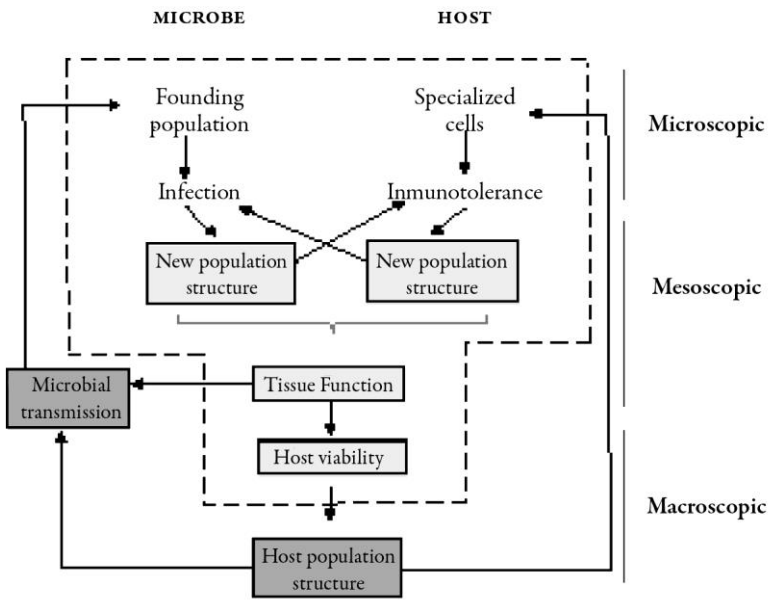

FIGURE 2. Nested equilibrium model. The dashed box represents those events that occur within the host. Adapted from Blaser & Kirschner (2007), p. 845, Fig. 2).

III. EQUILIBRIUM EXPLANATIONS AS STRUCTURAL AND NON-MECHANISTIC EXPLANATIONS

Blaser and Kirchner's NEM was developed to account for the persistence and the long-term character of certain human-microbe associations. Concretely, the authors seek to explain two paradoxes: first, why the association is not disrupted by the appearance of cheaters, i.e. entities that ben-

efit from the association without paying the costs; second, why the bacterial population is not entirely substituted by an intruder/external invader that deploys a different strategy. Only if those two phenomena are avoided, persistent host-bacterial associations can be successful. We will now argue that Blaser and Kirchner's NEM explains how those phenomena are avoided by appealing to mathematical, *but not causal*, properties, of host-microbial associations. In other words, we will argue that the alleged explanatory force of the NEM lies in the fact that: (i) it provides a mathematical structure in the form of a set of differential equations that together satisfy an ESS; (ii) that the nested nature of the ESSs makes the explanation of host-microbe persistence robust to any perturbation; (iii) that this is so because the properties of the ESS directly mirror the properties of the biological system in a non-causal way.

First of all, as shown in section II, Blaser and Kirschner's NEM consists in a series of differential equations that describe how the concentration of bacteria in different host tissues, their effector cells, their nutrient availability, the immunological response and their rate of transmission will change over time. These equations, as we explained, do not contain *a priori* any information about the persistence of the host microbe relationship. However, they provide information about how the different variables must be related to each other so that persistence obtains. Particularly, the equations measure the impact of host immunological response on bacterial colonization and, in doing so, allow determining the level at which host's response will abruptly disrupt colonization, as well as the levels at which bacterial inflammation will trigger a decrease in nutrient availability that in the end will disrupt colonization. And, in addition, they provide information about the way in which the solutions to these equations that guarantee the persistence of the symbiotic relation relate to: a) the rate of transmission of the symbiont (R_o), b) the viability of the host (tissue function and evolutionary advantages).

The set of equations can be resolved for a concrete host-symbiont system, and the evolution of the variables under study, as well as their interrelation, can be analysed. This will provide information about how they relate and how they are maintained constant, allowing predictions about empirical system[9]. However, notice that they would still provide no information about our *explanandum*, i.e. about what makes the host-microbe relationship persistent. To do so, the set of equations must be embedded in the framework of ESSs, i.e. it must model the biological situation as a non-cooperative game of two players, such that if any of the players (host, microbe) follows a unilateral strategy, the consequences

will be detrimental for the player that does so. That this is so can be seen by studying how changes in the equations that relate the concentration in nutrient availability, immune response, microbial concentration, etc. will relate to each other to make the system collapse if the change is unilateral. However, as we argued, the explanatory character of the equations comes exclusively from the possibility of embedding them in the framework of ESS. In other words, they are explanatory sound because it is possible to realize that no unilateral change that disrupts the system is possible without generating a chain reaction that either reverses the change or destroys the system. The ESS thus explains stability by ruling out two alternative scenarios: one where cheaters spread in the population, and another when an invader population entirely substitutes the actual one.

Second, the explanatory force of the ESS is reinforced in Blaser and Kirschner's NEM due to its nested nature. The nested nature of the equilibria works as a check and balances system which prevents that a disruption of the ESS at one of the levels (microscopic, mesoscopic and macroscopic) spreads across the other levels and destroys the host-microbe association. Let us explain this with an example: take the case of a disruption at the mesoscale that substitutes the microbe population for an invader. As we are at the mesoscale, the invader will disrupt tissue function in its own benefit, e.g. growing more than what the original microbial population would have grown, while at the same time escaping from the barriers of the immunological system. This type of change, totally beneficial for the bacteria at the mesoscale, would trigger two responses: First, a response at the macroscale that would be immediately detrimental for the bacteria. At this level, host viability, which is affected by the tissue function, will be reduced and, as a consequence, bacterial transmission will substantially decrease in relation to the transmission of those bacteria that cause no damage in tissue function. Secondly, at the microscale, where the invader population will not have generated immunotolerance, the invader population will be systematically blocked by the specialized immunological cells, especially the cells of the adaptive immune system. Furthermore, it is expected that the host will reduce nutrient availability, so that it affects in the long-run the intruders' population structure. Remember, as we said in section II, that the key of the ESS is that no player that changes its strategy unilaterally will be better. In this situation, even if the "player" might be better in one particular scale (mesoscale), the same will not be true for the other scales, and thus no possibility for invasion exists[10].

Third, and more concretely about the nature of ESS, we believe that Blaser and Kirschner's NEM, as any explanation that appeals to the existence of an ESS, explains the stability of host-microbe persistent associations in a non-causal way. Let us argue why we believe this to be so.

1) Blaser and Kirschner's NEM appeals to general properties of ESSs, and they make their model explanatory in virtue of the equivalence between the theoretical ESSs framework and the general properties of persistence host-symbiont associations. The strategy is the general strategy of Huneman's structural explanations: first, build a system *S'* whose properties match the properties of the real system *S* whose behaviour you aim to track. Second, study the behaviour of *S'* and attribute its properties to *S*. In Blaser and Kirschner's NEM, the strategy is applied as follows: first, build the ESS model for host-microbe persistent associations, as a case of a non-cooperative game of two players; second, study the behaviour of the ESS model, i.e. why the existence of an ESS, as the optimal solution for both players (Nash equilibrium), excludes the possibility of cheaters and invasive populations; third, attribute the properties of the ESS model to the empirical phenomenon, i.e. to empirical cases of host-microbe persistent associations. Notice that in this schema the explanatory force comes because the mathematical system that is built, in this case an equilibrium model, behaves in a certain way that (allegedly) is the way in which the empirical system will behave. But, importantly, it is irrelevant how the empirical phenomenon causally realizes the properties that it is attributed. And this is so in a double sense: on the one hand, because the NEM neither mention, nor needs to mention the specific species that interact to generate the ESS; on the other, because the causal connections between the entities (*if any*) are epistemologically irrelevant for the explanation of the phenomenon.

2) Despite the highly problematic way of identifying interlevel causal relations in a multilevel mechanism, as Craver and Bechtel (2007) explain, one could still try to appeal to Woodward's interventionist strategy to identify the supposed causes explaining the persistence of host-microbe associations. However, we believe NEM rules out the possibility of generating or even heuristically imagining any intervention *à la* Woodward, thus contradicting Kuorikoski and Potochnick's interpretation of equilibrium explanations. Let us explore this via an example. Recall that the *explanandum* is the phenomenon of persistence host-microbe associations. How would an intervention look like in Blaser and Kirchner's NEM? The only possibility would be to generate a situation such that the ESS disappears. However, no possible intervention is imaginable without destroying the system. Or,

in other words, any imaginable intervention that would make host-microbe associations non-persistent would directly change the system we are trying to explain, and thus the information it will provide will turn out to be irrelevant to account for the phenomenon. Recall the structure of ESS (Table 2). The only possibility of imagining a significant intervention would be via a change in the expected payoffs for the actions of each player. However, this intervention would not give any relevant information about why the association is stable in certain circumstance, because it would directly shift the focus of attention towards a new system, namely, one where there is not an ESS. Or, in other words, a causal explanation would consist in saying that the ESS is explanatory because if there were not an ESS the host-microbe association would not be stable. But this kind of reasoning is uninformative and, in our view, unexplanatory. The structural interpretation *à la Huneman*, on the contrary, offers a plausible account of how Blaser and Kirschner's NEM gains its explanatory force.

More importantly, the nested nature of the model, far from moving its explanatory force in a causal-mechanistic direction, generates the opposite effect. It just makes any possible intervention less imaginable. Because even if one causal intervention could be imagined for one specific level, how would it possibly work, if its effects would be cancelled out due to the existence of ESSs in the other levels? Or, in other words, how is it possible to imagine an intervention that causally escapes the inter-level connection? This connection is just a property of any host-microbe persistent association, and the explanatory power of the nestedness resides, precisely, in its possibility to cancel out the effect of every possible intervention. Therefore, we argue, a causal interpretation of the explanatory power of Blaser and Kirschner's NEM is not possible, since it would simply make the explanatory force of the model completely mysterious.

Of course, one might agree with what we just said, and still believe that our argument does not rule out the fact that the most appropriate interpretation of the explanatory force of Blaser and Kirchner's NEM is indeed causal. For instance, Blaser and Kirschner explicitly argue that specific host-microbe associations (human-*H. pylori*, human-*Salmonella typhi*, etc.) are "not necessarily forever stable" [(2007), p. 843], as obviously context (environment) matters, and in a changing context (environment) it is possible that concrete associations go selected against, simply because the environment selects against that coevolved system [see Díaz (2015); Suárez & Triviño (2019)]. In this context, it is possible to investigate the causes that made the system collapse, and if this is so, then the same must

be true for the cases in which the association is persistent. Nonetheless, we disagree, because that will entail changing the *explanandum* in two senses: first, making it specific to particular species; second, explaining the disruption of the persistence, instead of the persistence itself. And remember that our original *explanandum* was why some host-microbe associations are persistent, and the cases to rule out are the cases of cheaters and invasive populations. In our view, their model should be interpreted counterfactually: if a host-microbe association is persistent throughout the host's life cycle and evolutionarily long-term, then it will satisfy the conditions of the NEM reached through an ESS. And this situation will be so irrespectively of the species that interact, and thus irrespective of the causal-mechanisms that host and microbe could have developed to reach that equilibrium. As in the case of the ice vendors (section I), where the psychological mechanisms that have driven the vendors to put their stalls in the middle of the beach are explanatorily irrelevant to understand why their stalls are there, in the case of persistent associations causal-mechanistic details are simply superfluous. One can perfectly omit all those details and the explanation would still be epistemically sound.

Alternatively, an enumeration of the causes (if any) that would determine whether a concrete host-microbe association is stable will be irrelevant to explain its persistence if it is not conceived as a consequence of an ESS. This is because it would still be possible to imagine the existence of cheaters or invasive populations that deploy the same causal-mechanistic "machinery" to escape e.g. immunitary controls, without paying the cost of the symbiotic association. However, as we explained, because the host-microbe association constitutes a nested ESS, both the cheater and the invader population will end up disappearing from the population, just because the host-microbe persistent system has the structure that appears in the mathematical formulation of ESSs. Importantly, we are not here saying that Blaser and Kirchner's NEM rules out the possibility of telling a causal story of why concrete host-microbe associations are, sometimes, persistent, although some story about how to speak about interlevel causation should be provided.[11] Furthermore, we believe that such causal stories *could* be told to explain specific host-microbe associations, even when these must be complemented with the appeal to ESSs. Our point is rather *epistemological*: causal stories that seek to explain the existence of persistent host-microbe associations are neither required, nor explanatory in themselves. The element that provides the explanatory strength in equilibrium explanations is purely structural (in Huneman's terms), and it is connected with the possibility of accounting for the existence of an equilibrium (in Blaser and Kirschner's NEM, a nested ESS).

IV. Conclusion

In this paper, we have examined the explanatory force of equilibrium explanations, and have studied whether the explanatory force of equilibrium explanations can be better justified by applying the causal-mechanistic model of scientific explanation, or Huneman's structural model. Concretely, we have examined the role that mathematical vs. causal properties play in the explanation of the stability of persistent long-term host-microbe associations. Explaining the stability of this type of associations is paradoxical, as it requires explaining two facts: first, the absence of cheaters; second, the impossibility of the population being substituted by an intruder population. We have used Blaser and Kirschner's NEM to illustrate that the explanation of host-microbe persistent associations does not seem to be causal, but structural, relying solely on the non-causal mathematical properties of the association to explain its long-term persistence [Huneman (2018a), (2018b)]. We have argued that Blaser and Kirschner's NEM is explanatory of the long-term persistence of host-microbe associations because (i) it provides a mathematical structure in the form of a set of differential equations that together satisfy an ESS; (ii) that the nested nature of the ESSs makes the explanation of host-microbe persistence robust to any perturbation; (iii) that this is so because the properties of the ESS directly mirror the properties of the biological system in a non-causal way. In this vein, our case study shows how equilibrium explanations, even if nested, gain their explanatory force from the mathematical structure that describes the system, instead of from the causal interactions among its components. Our analysis supports two theses: first, that equilibrium explanations, even if nested (in a hierarchical setting), are structural rather than causal-mechanistic; second, that causality, even if necessary in some explanations, is not a universally necessary requirement of every scientific explanation.

Logos – Research Group in Analytic Philosophy / Barcelona Institute of Analytic Philosophy
Department of Philosophy
University of Barcelona
C/Montalegre nº6, 08001. Barcelona, Spain
E-mail: javier.suarez@ub.edu
E-mail: roger.deulofeu@gmail.com

ACKNOWLEDGMENTS

A previous version of this paper was presented in the meeting 'Process epistemology: A workshop with Bill Bechtel', University of Exeter, May 2017. The authors want to thank all the participants for their feedback, and especially to Bill Bechtel for a wonderful discussion. José Díez, John Dupré, Philippe Huneman, Thomas Pradeu, Johannes Findl and two anonymous reviewers are acknowledged for their comments on earlier versions of the paper. The following institutions are formally acknowledged: Javier Suárez and Roger Deulofeu, Spanish Ministry of Economy and Competitiveness (FFI2016-76799-P); Roger Deulofeu, Spanish Ministry of Economy and Competitiveness (BES-2013-063239); Javier Suárez, Spanish Ministry of Education (FFU16/02570).

NOTES

[1] The commitment to a causal view of the world does not entail either a physical reductionism [as in Salmon (1984)] or an "ontic" interpretation of scientific explanation [as in Craver (2014)]. Cf. Glennan (2002), Bechtel & Abrahamsen (2005), for a model-based interpretation of mechanisms.

[2] There are other cases where equilibrium models have been used to explain the stability of biological associations [Baalen & Jansen (2001); Selosse *et al.* (2006)]. We have chosen to analyse Blaser & Kischner's NEM for its generality, and because it is a case of equilibrium explanation generally accepted among biologists. Nonetheless, our conclusions also apply to these cases. Thanks to Philippe Huneman for pointing this fact to us.

[3] Following Brigandt (2013), we consider that an element of an *explanans* is explanatory relevant if and only if removing it from the explanation entails that the *explanandum* does not follow, and it's explanatory irrelevant otherwise [(2013), p. 480].

[4] "Microbiota" refers to "[t]he assemblage of microorganisms present in a defined environment", and "microbiome" is used to denote "the entire habitat, including the microorganisms (bacteria, archaea, lower and higher eukaryotes, and viruses), their genomes (i.e., genes), and the surrounding environmental conditions" in a given environment [Marchesi & Ravel (2015), p. 1]. For the purposes of this paper, we will not distinguish the two concepts, and they will be used to refer only to the community of microorganisms present in a given environment.

[5] In biology, persistent infection refers to lifelong associations between a host and some species of microbes that do not necessarily harm the host, although they might do it in the long-term. The term should not be confused with its medical use, where "infection" is usually employed in reference to pathogens, or disease-causative agents.

[6] Their model is in principle developed exclusively for pair associations, between one host and one microorganism.

[7] Those different levels have both a temporal and a scale correlation: the macroscale refers to the evolutionary time, the mesoscale refers to organismal development and the microscale refers to the interactions among different cell types.

[8] Since our purpose is only to illustrate the main features of the model and their relation to Blaser and Kirschner's explanation, for a matter of simplicity we only introduce two of the equations.

[9] Information about the values that the variables must take for a concrete (empirically real) host-microbe association, if the association is known to be stable.

[10] It exists, but if and only if the intruder changes the situation *in the three scales*. That is precisely the nature of the nested model.

[11] See Craver & Bechtel (2007) for a proposal.

REFERENCES

BATTERMAN, R.W., and C. C. RICE (2014), "Minimal Model Explanations"; *Philosophy of Science* 81. 3, pp. 349-376.

BECHTEL, W., and A. ABRAHAMSEN (2005), "Explanation: A Mechanist Alternative"; *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2), pp. 421–441.

BLASER, M. J. (2006), "Who Are We? Indigenous Microbes and the Ecology of Human Diseases"; *EMBO Rep* 7, pp. 956–960.

BLASER, M. J. and J. ATHERTON (2004), "*Helicobacter pylori* Persistence: Biology and Disease"; *J. Clin. Invest*. 113, pp. 321–333.

BLASER M. J., and D. KIRSCHNER (1999), "Dynamics of *Helicobacter pylori* Colonization in Relation to the Host Response"; *Proc Nat Acad Sci* 96, pp- 8359-8364.

— (2007), "The Equilibria that Allow Bacterial Persistence in Human Hosts"; *Nature* 449, pp. 843–849.

BRIGANDT, J. (2013), "Systems Biology and the Integration of Mechanistic Explanation and Mathematical Explanation"; *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44, pp. 477–492.

BRIGANDT, J., S. GREEN, and M. A. O'MALLEY (2017)," Systems Biology and Mechanistic Explanation"; in S. Glennan and P. Illari (eds.) *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, London: Routledge (chapter 27).

CHIU, L., and S. F. GILBERT (2015), "The Birth of the Holobiont: Multi-species Birthing Through Mutual Scaffolding and Niche Construction"; *Biosemiotics* 8 (2), pp. 191–210.

CRAVER, C. F. (2007), *Explaining the Brain*; New York: Clarendon Press.

— (2014), "The Ontic Account of Scientific Explanation"; in Kaiser, M. I., Scholz, Plenge, R. D. Hüttemann, A. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Springer Verlag. pp. 27–52.

CRAVER, C. F., and W. BECHTEL (2007.), "Top-Down Causation Without Top-Down Causes"; *Biology & Philosophy* 22(4), pp. 547-563.

CRAVER, C. F., and L. DARDEN (2013), *In search for Mechanisms: Discovery Across the Life sciences*; Chicago: University of Chicago Press.

DEULOFEU, R. and J. SUÁREZ (2018), "When Mechanisms Are Not Enough: The Origin of Eukaryotes and Scientific Explanation"; in Christian A., Hommen D., Retzlaff N., Schurz G. (eds) *Philosophy of Science*. European Studies in Philosophy of Science, vol 9. Springer, Cham.

DEULOFEU, R, J. SUÁREZ and A. PÉREZ-CERVERA (2019), "Explaining the Behaviour of Random Ecological Networks: The Stability of the Microbiome as a Case of Integrative Pluralism"; *Synthese. https://doi.org/10.1007/s11229-019-02187-9.*

DÍAZ, J. S. (2015), "El Mecanismo Evolutivo de Margulis y los Niveles de Selección"; *Contrastes* XX (1), pp. 7-26.

DOUGLAS A. E. and J. H. WERREN (2016), "Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts"; *mBio* 7 (2), e02099-15.

EASLEY, D. and KLEINBERG, J. (2010), *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*; Cambridge University Press.

EBERL, G. (2016), "Immunity by Equilibrium"; *Nat. Rev. Immunol.* 16, pp. 524-532.

GINTIS, H. (2000), *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Behavior*.; Princeton University Press.

GLENNAN, S. (2002), Rethinking Mechanistic Explanation; *Philosophy of Science* 69 (S3), pp. S342–S353.

GLENNAN, S. and ILLARI, P. (Eds.) (2017), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, Taylor and Francis.

GREEN, S. (2016), *Philosophy of System Biology*; Dordrecht: Springer.

— (2017), "Philosophy of Systems and Synthetic Biology"; *The Stanford Encyclopedia of Philosophy* (Edition Spring 2019), Edward N. Zalta (ed.), URL = <*https://plato.stanford.edu/archives/spr2019/entries/systems-synthetic-biology/*>.

GREEN, S. and JONES, N. (2016), "Constraint-Based Reasoning for Search and Explanation Strategies for Understanding Variation and Patterns in biology"; *Dialectica* (70)3, pp. 343-374.

HUNEMAN, P. (2010), "Topological Explanations and Robustness in Biological Sciences"; *Synthese* 177, pp. 213–245.

— (2018a), "Outlines of a Theory of Structural Explanation"; *Philosophical Studies* 175 (3), pp. 665–702.

— (2018b), "Diversifying the Picture of Explanations in Biological Sciences: Ways of Combining Topology with Mechanisms"; *Synthese* 195, pp. 115–146.

— (2018c); "Realizability and the Varieties of Explanation"; *Studies in History and Philosophy of Science. <https://doi.org/10.1016/j.shpsa.2018.01.004>*

HUTTENHOWER C, GEVERS D, KNIGHT R, CREAS HH, *et al.* (2012) "Structure, Function and Diversity of the Healthy Human Microbiome"; *Nature* 486, pp. 207–214.

ISSAD, T., and C. MALATERRE (2015.), "Are Dynamic Mechanistic Explanations Still Mechanistic?"; in P. A. Braillard and C. Malaterre (eds.) *Explanation in*

Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences. Dordrecht: Springer, pp. 265–292.

JONES, N. (2014), "Bowtie Structures, Pathway Diagrams, and Topological Explanations"; *Erkenntnis* 79 (5), pp. 1135–1155.

KAPLAN, D. M. and C. F. CRAVER (2011), "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience. A Mechanistic Perspective"; *Philosophy of Science* 78.4, pp. 601-627.

KITANO, H., and K. ODA (2006), "Robustness Trade-Offs and Host-Microbial Symbiosis in the Immune System"; *Molecular Systems Biology* 2, pp. 1–10.

KOSTIĆ, D. (2018), "The Topological Realization"; *Synthese*, 195(1), pp. 79-98.

— (2019), "Minimal Structure Explanations, Scientific Understanding and Explanatory Depth", *Perspectives on Science*, 27 (1), pp. 48-67.

KRICKEL, B. (2018), *The Mechanical World: The Metaphysical Commitments of the New Mechanistic Approach*; (Vol. 13). Springer.

KUORIKOSKI, J. (2007), "Explaining with Equilibria"; in Persson, J., and Ylikoski, P. (Eds.), *Rethinking explanation*; Springer, Dordrecht, pp. 149-162.

LOZUPONE, C. J. I. STOMBAUGH, J. I. GORDON, J. K. JANSSON, and R. KNIGHT (2012), "Diversity, Stability and Resilience of The Human Gut Microbiota"; *Nature* 489 (7415), pp. 220–230.

MACHAMER, P., DARDEN, L., and C.F. CRAVER (2000), "Thinking About Mechanisms"; *Philosophy of science*, 67(1), pp. 1-25.

MARCHESI, J. R., and J. RAVEL. (2015), "The Vocabulary of the Microbiome Research: A Proposal"; *Microbiome* 3, p. 31.

MORAN, N., and D. B. SLOAN (2015), "The Hologenome Concept: Helpful or Hollow?"; *PLoSBiol* 13 (12), e1002311.

MORENO A., and J. SUÁREZ (submitted), "Plurality of Explanatory Strategies in Biology: Mechanisms and Networks".

NASH, J. F. (1950a), "The Bargaining Problem"; *Econometrica: Journal of the Econometric Society* 18 (2), pp. 155-162.

— (1950b), "Equilibrium Points in N-Person Games"; *Proceedings of the National Academy of Sciences* 36 (1), pp. 48-49.

NICHOLSON, D.J. (2012), "The Concept of Mechanism in Biology"; *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1), pp. 152–163.

POTOCHNIK, A. (2015) "Causal Patterns and Adequate Explanations"; *Philosophical Studies*, 172(5), pp. 1163-1182.

PRADEU, T., JAEGER, S., and E. VIVIER (2013), "The Speed of Change: Towards a Discontinuity Theory of Immunity?"; *Nature Reviews Immunology*, 13(10), p. 764.

PRADEU, T., and E. VIVIER (2016), "The Discontinuity Theory of Immunity"; *Sci. Immunol.* 1 (1): aag0479.

RICE, C. (2015), "Moving Beyond Causes: Optimality Models and Scientific Explanation"; *Noûs* 49.3 pp. 589-615.

ROSENBERG E and I ZILBER-ROSENBERG (2014), *The Hologenome Concept*. London, Springer.

— (2016) "Microbes Drive Evolution of Animals and Plants: The Hologenome Concept"; *mBio* 7 (2): e01395-15.

SALMON W. (1984), *Scientific Explanation and the Causal Structure of The World*; Princeton: Princeton University Press.

SELOSSE, M. A., RICHARD, F., HE, X., and S.W. SIMARD (2006), "Mycorrhizal Networks: des Liaisons Dangereuses?"; *Trends in Ecology and Evolution* 21 (11), pp. 621-628.

SMITH, J. M. (1974), "The Theory of Games and the Evolution of Animal Conflicts"; *Journal of Theoretical Biology*, 47(1), pp. 209-221.

SMITH, J. M., and G. R. PRICE (1973), "The Logic of Animal Conflict"; *Nature*, 246(5427), p. 15.

SOBER, E. (1983), "Equilibrium Explanation"; *Philosophical Studies* 43.2, pp. 201-210.

SUÁREZ, J. (2018). "The Importance of Symbiosis in Philosophy of Biology: An Analysis of the Current Debate on Biological Individuality and its Historical Roots"; *Symbiosis* 76(2) pp. 77-96.

SUÁREZ, J. and V. TRIVIÑO (2019), "A Metaphysical Approach to Holobiont individuality: Holobionts as Emergent Individuals"; *Quaderns de Filosofia* 6(1), pp. 59-76.

VAN BAALEN, M. and V.A. JANSEN (2001), "Dangerous Liaisons: The Ecology of Private Interest and Common Good": *Oikos*, 95(2), 211-224.

WOODWARD, J. (2003), *Making Things Happen: A Theory of Causal Explanation*; New York: Oxford University Press.

— (2013), "Mechanistic Explanation: Its Scope and Limits"; *Aristotelian Society Supplementary Volume* 87 (1), pp. 39–65.

Synthese
https://doi.org/10.1007/s11229-019-02187-9

BIOBEHAVIOUR

Explaining the behaviour of random ecological networks: the stability of the microbiome as a case of integrative pluralism

Roger Deulofeu[1] · **Javier Suárez**[1,2] · **Alberto Pérez-Cervera**[3]

Received: 12 June 2018 / Accepted: 20 March 2019

Abstract

Explaining the behaviour of ecosystems is one of the key challenges for the biological sciences. Since 2000, new-mechanism has been the main model to account for the nature of scientific explanation in biology. The universality of the new-mechanist view in biology has been however put into question due to the existence of explanations that account for some biological phenomena in terms of their mathematical properties (mathematical explanations). Supporters of mathematical explanation have argued that the explanation of the behaviour of ecosystems is usually provided in terms of their mathematical properties, and not in mechanistic terms. They have intensively studied the explanation of the properties of ecosystems that behave following the rules of a non-random network. However, no attention has been devoted to the study of the nature of the explanation in those that form a random network. In this paper, we cover that gap by analysing the explanation of the stability behaviour of the microbiome recently elaborated by Coyte and colleagues, to determine whether it fits with the model of explanation suggested by the new-mechanists or by the defenders of mathematical explanation. Our analysis of this case study supports three theses: (1) that the explanation is not given solely in terms of mechanisms, as the new-mechanists understand the concept; (2) that the mathematical properties that describe the system play an essential explanatory role, but they do not exhaust the explanation; (3) that a non-previously identified appeal to the type of interactions that the entities in the network can exhibit, as well as their abundance, is also necessary for Coyte and colleagues' account to be fully explanatory. From the combination of these three theses we argue for the necessity of an integrative pluralist view of the nature of behaviour explanation when this is given by appealing to the existence of a random network.

The paper is the result of the discussion among the three authors, who actively collaborated in the development of all the ideas. JS conceived and structured it. RD and APC wrote Section 3. RD and JS wrote the philosophical analysis.

✉ Roger Deulofeu
Roger.deulofeu@gmail.com

Extended author information available on the last page of the article

Published online: 03 May 2019

Keywords Scientific explanation · Mechanism · Mathematical explanation · Behaviour explanation · Integrative pluralism · Random network

1 Introduction

Explaining the behaviour of ecosystems is one of the key challenges for biologists: why ecosystems have the properties they have, which conditions make them exhibit a stable behaviour, how they react to perturbations, etc. have been some of the most debated questions among behavioural ecologists. Almost since the original publication of the "new-mechanist" *manifesto* (Machamer et al. 2000), the idea that explanation in biology proceeds by discovering mechanisms has been extensively accepted among philosophers of science (Bechtel and Richardson 1993; Glennan 2002; Bechtel and Abrahamsen 2005; Craver and Darden 2005, 2013; Craver 2007). Drawing upon the findings in neuroscience and molecular biology, new-mechanists reelaborate causalism (Salmon 1984; Woodward 2003) and argue that to explain a phenomenon (*explanandum*) consists in describing its causes by providing a mechanism (*explanans*) responsible for the phenomenon coming about. According to the new-mechanist interpretation of explanation, a phenomenon can be considered fully explained only once a mechanism is provided and the way its components causally interact to produce the phenomenon is specified.

The universality of mechanistic explanations in biology and neuroscience as originally formulated in the new-mechanist *manifesto* has however been questioned on different grounds. A very popular "non"-mechanist view holds that some explanations in biology are at least *partially* given by appeal to the mathematical properties of the systems under investigation. Some of those who defend the necessity of introducing mathematical modelling in some biological explanations have emphasized the continuity of these explanation with mechanistic explanations, developing the concept of *dynamic mechanistic explanations* (Bechtel and Abrahamsen 2010; Bechtel 2011; Brigandt 2013a, b, 2015), whereas others have highlighted the differences between both accounts, arguing that dynamic mechanistic explanation are indeed no-mechanistic (Issad and Malaterre 2015). A still more radical interpretation of this "non"-mechanist approach holds that some explanations in biology are given exclusively in mathematical terms, with no reference to any type of causal-mechanistic information in the *explanans* (Sober 1983; Huneman 2010; Lange 2013; Jones 2014). This last kind of explanation might be called *mathematical explanation*, and it is usually presented in complete opposition to new-mechanist models of explanation.

One of the most recent topics of attention of defenders of the last approach has been the study of the explanation of the behaviour of ecosystems. Some of the most salient behavioural properties of ecosystems (robustness, resilience, stability), they argue, are explained using the tools provided by network analysis (Xia 2010; Rathkopf 2018). The use of these mathematical tools to explain how some biological systems behave in the long-term is very abundant in ecology (e.g. the diversity-stability debate) due to the complexity of ecological systems. It has recently been argued that explanations of the behaviour of ecosystems are given exclusively on the basis of the mathematical properties of the networks that are used to represent them (Huneman 2010, 2018a, b,

c). The study of behaviour explanations that appeal to network analysis in ecology has focused exclusively on the study of ecosystems that form *non-random networks*, though. However, as far as we know, the explanatory features of behaviour explanations in ecosystems that organize forming *random networks* have received no attention among scholars. A network is said to be non-random when the aggregation of the elements that interact in the network lead to a concrete topological realization (small world, scale-free, etc.) with some a priori known properties. A random network, on the contrary, is characterized for lacking a known topological realization, and thus the properties of a random network cannot be known a priori, but have to be mathematically discovered. We suspect that the difference between systems that organise according to a random network and systems that organise according to a non-random network get their explanatory force from different sources. In this paper, we aim to test that intuition by studying the explanatory features of the models that explain the ecological behaviour of a random network—the human microbiome.

A microbiome is a collection of microorganisms (bacteria, viruses, fungi, etc.) that reside in a concrete environment (Marchesi and Ravel 2015). In case of humans, our microbiome is believed to contain hundreds of species (Human Microbiome Project Consortium 2012). Furthermore, human's gut microbiome is known for its ecological stability behaviour: even if it varies quite a lot from one person to another, the gut microbiome is believed to be very stable for one single individual, who tends to carry the same species of microbes for a long period (Dethlefsen and Relman 2011; Faith et al. 2013). Recently, Coyte et al. (2015) and Foster et al. (2017) have elaborated a model to explain why the human gut microbiome exhibits a stable behaviour over long periods of time. They found out that, contrary to a usual assumption in evolutionary biology, competition, and not cooperation, is the key factor explaining this stable behaviour. Here, we analyse how Coyte and colleagues explain the stability of the microbiome by appealing to linear stability analysis. This case shares several elements with other cases of behaviour explanation in terms of networks, although it also differs in some features that led to some differences regarding the nature of behaviour explanation and that we aim at clarifying.

In Sect. 2 we introduce the notions of mechanistic and mathematical explanation, and frame the choice of our case study in the context of the mechanistic versus mathematical debate. In Sect. 3 we present Coyte and colleagues' explanation of the stability behaviour of the human microbiome. In Sects. 4 and 5 we discuss the consequences of the case study for the understanding of scientific explanation. In Sect. 4 we argue that even if the explanation in our case study displays a model of mechanism, it does not specify a proper causal story to account for the *explanandum*, thus suggesting that mechanisms, as understood by new-mechanists, do not play the explanatory role. In Sect. 5 we argue that the explanatory force in Coyte and colleagues' account comes from the display of a mathematical model of the behaviour of the microbiome, as it is provided by their linear stability analysis of the network that the microbiome instantiates. However, in contrast with some recent analysis of behaviour explanations of ecosystems in terms of networks, we argue that insofar as an essential component of the explanatory force of Coyte et al.'s model is their appeal to the different interaction types within the ecological communities that might exist in the microbiome (cooperative, exploitative, competitive), and thus to the establishment of a range of topologies—and

not exclusively to a concrete topology—their explanations differs from purely mathematical explanations. We further argue that the appeal to this element is exclusive of random networks. In Sect. 6 we explore the pluralistic consequences of our case study for the analysis of behaviour explanation, suggesting that it supports the general appeal to integrative pluralism. Finally, we present our concluding remarks.

2 Two models of explanation in biology: mechanistic versus mathematical

In its most basic meaning, to explain a phenomenon consists in giving the reasons why the phenomenon obtains in a concrete system due to its behaviour. Those reasons may be specified by appealing to laws of nature—deductive-nomological theories of explanation—or by appealing to the causes that are responsible for the phenomenon. In the first case, it is usually assumed that to explain a phenomenon is to provide an argument so that the phenomenon (*explanandum*) logically follows from the laws of nature that regulate the behaviour of the system where it obtains, given a concrete set of initial conditions (*explanans*) (Díez 2014; Alleva et al. 2017). In the second case, the phenomenon is embedded in a causal network in the world, in a way such that the phenomenon (*explanandum*) is a causal consequence of the behaviour of the system (*explanans*) (Woodward 2017).

New-mechanist theories of scientific explanation are of this last kind. They assume that to explain a phenomenon consist in citing its causes by providing a mechanism that specifies how those causes produce the phenomenon. The definition of "mechanism" is different in differing accounts (Nicholson 2012; Deulofeu and Suárez 2018), although most new-mechanists share a similar conception: a mechanism consists in a set of *entities* with a concrete spatial *organization* plus a set of *activities* governing the behaviour of those entities (*model of the mechanism*). The phenomenon to explain, new-mechanists argue, *causally* obtains as a consequence of the activities of the entities (*causal story*) (e.g. Machamer et al. 2000; Bechtel and Abrahamsen 2005; Craver and Darden 2005 2013; Craver 2007; etc.). For new-mechanists, the presence of a *model of the mechanism* and of a *causal story* is necessary and sufficient for having a mechanistic explanation (Issad and Malaterre 2015: p. 270).

New-mechanists generally accept a hierarchical view of mechanisms, thus not confining themselves to the narrow approaches of previous causalists (Salmon 1984). Furthermore, they usually neglect the capacity of mathematical models *alone* to explain any phenomenon. In a well-known paper, Kaplan and Craver argued that "the [mathematical] generalizations are explanatory because they describe the causal relationships that produce, underlie, or maintain the *explanandum* phenomenon" (2011: p. 612) and insisted that:

> In successful explanatory models in cognitive and systems neuroscience (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifi-

able) causal relations among the components of the target mechanism. (2011: p. 611).

Mathematics, thus, can play an explanatory role for new-mechanists only if it captures *a causal relationship* among the entities that are posited in the mechanism. Otherwise, they are merely "phenomenological models", which represent the reality and allow predictions without really explaining why those predictions obtain (e.g. Kepler's laws, Snell's laws, etc.) (Díez 2014).

Radically contrasting with this last view, some people have vindicated a more substantial role for mathematics in biological explanations by highlighting the importance of mathematical properties for explaining the features of some biological systems. The explanations that rely on the mathematical properties of the system to explain a phenomenon have been called "mathematical explanations" (Baker 2015), or "structural explanations" (Huneman 2018a).[1] They have been defined as follows:

> Family of explanations for which the mathematical tools used in the description of an explanandum system belong to a mathematical structure whose properties are directly explanatory of some aspects of the system (such as equilibria, behaviour, limit regime, asymptotic behaviour, etc.) (…) They explain by accounting for the explananda through pinpointing structural relations that are mathematical relations of some sort. (Huneman 2018a: p. 695)

The mathematical properties that appear in a mathematical explanation might be of different types, and they could be used to explain different kinds biological questions. They might consist in: the application of an arithmetic theorem to explain the life cycle of some species (Baker 2005, 2009, 2015); the establishment of one or more points of equilibrium to explain a tendency in a population (Sober 1983; Kuorikoski 2007; Rice 2012, 2015; Suárez and Deulofeu, unpublished manuscript); the application of statistics to explain certain evolutionary patterns in a population (Walsh 2015); the discovery of a concrete topology that explains the behaviour of a complex system (Huneman 2010; Jones 2014); the use of matrix calculus to explain the processes that regulate some physiological states (Issad and Malaterre 2015); etc.[2]

Not every substantial use of mathematics in biological explanations needs to be in principle completely opposed to every element of the new-mechanistic account of explanation, though. The explanation of some cyclical biological processes such as the circadian rhythms has been argued to constitute an extension of mechanistic explanation, namely a *dynamic mechanistic explanation* (Bechtel and Abrahamsen 2010, 2011; Brigandt 2013a, b, 2015). In general, an explanation is considered a dynamic mechanistic explanation in virtue of making use of some mathematical model in its *explanans* that: (1) is essential to account for the *explanandum*, (2) replaces the role that new-mechanists attribute to the causal story, without being itself a causal

[1] To refer to mathematical explanations as "structural explanations" might be confusing, since the later could be interpreted as special cases of the former, as one reviewer has correctly suggested. However, the way in which Huneman (2018a) describes them, as well as the family of explanations that he includes under the umbrella of "structural explanations" makes clear that the two are synonymous. For purposes of clarity, however, we will refer to this family of explanations as "mathematical explanations".

[2] We take all the aforementioned properties to be different types of mathematical properties.

story *stricto* sensu, (3) is combined with a model of mechanism, i.e. a set of entities and activities plus their organization. Nonetheless, it is precisely because of the lack of a causal story that some of these explanations entail that they have been argued to be "anti"-mechanistic, or at least far away from the core elements that new-mechanists consider necessary and sufficient to formulate an adequate scientific explanation (Issad and Malaterre 2015).

In still more extreme cases, however, some explanations in biology have been argued to be even more substantially mathematical, abstracting away also from the model of the mechanism, and being explanatory of the biological phenomenon *exclusively* in virtue of the mathematical properties of the *explanans*, whatever their type (Sober 1983; Baker 2005; Huneman 2010; Jones 2014; cf. Kuorikoski 2007; Potochnik 2015). The analysis of behaviour explanations in biological systems that present a network structure (e.g. ecosystems, immunological systems, etc.) has been argued to follow this pattern of mathematical explanation. The behaviour of this type of systems is usually explained in two steps: (1) the system is attributed a concrete network topology, which provides the mathematical properties of the system; (2) the properties of this topology are studied and then its behaviour is attributed to the biological system, explaining why the system behaves how it does. Because of the reliance of the last type of mathematical explanation on the topological properties of the networks, it has been called *topological explanation* (Huneman 2010, 2018b; Jones 2014; Brigandt et al. 2017).

Those who have studied topological explanation more attentively have made two points: First, that the appeal to the topological properties of the mathematical structure alone (its graph structure, or its network motifs, for instance) is *sufficient* to explain some of the properties of the biological system that the structure is believed to represent, irrespectively of the entities and the activities of the entities that realize those systems (see also Huneman 2018c); second, that the addition of any mechanistic details, instead of making the explanation of the properties more precise, obfuscates the question and turns out to be deeply irrelevant for the embedding of the *explanandum*. This second point might be explicated as follows: in sharp contrast with dynamic mechanistic explanations, in topological explanations, neither the model of mechanism (nature of the entities, nature of the activities), nor the causal story are explanatory relevant. All that matters in the explanation is that the topology is provided and that it gives information about the organization of the system.

One of the fields where the appeal to topological properties to explain biological phenomena has proven more fruitful is in the diversity-stability debate in ecology (McCann 2000; Nikisianis and Stamou 2016). In that context, the aim of ecologists is to elaborate network models that represent the relationships among the biotic members of an ecological community with the aim of inferring some general features about its behavioural patterns. To do so, once the network model is elaborated and linked to behavioural properties of the ecological community which is being studied, ecologists analyse the global properties of the network—e.g. how it will respond to a perturbation, to an increasing/decreasing number of nodes, to an increase in the number of connections, etc., and then attribute the exact same properties to the ecological community that the network is believed to capture. The driving question in the diversity-stability debate is whether increasing the number of species in the community would make the

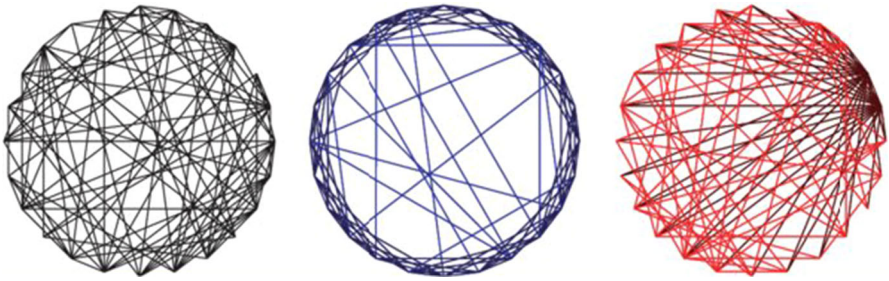

Fig. 1 This figure exemplifies different types of networks. The one on the left is a random network. The one in the middle is a small world network. The one of the right is a scale free network. (From Sporns et al. 2004: p. 419)

community ecologically stable[3] and, if so, under which conditions. Some recent discussion surrounding the diversity-stability hypothesis have tried to unravel how some ecological communities will react to the loss of some species for communities where the connections between the nodes are non-random (Solé and Montoya 2001). In the context of network theory, a network is non-random if it instantiates a particular topology, e.g. scale-free networks, small worlds (Fig. 1). Because the topological properties of theses types of networks are known, and both small worlds and scale-free networks are known to be highly stable to the elimination of some of their nodes (Montoya and Solé 2002), it is *enough* for ecologists to prove that a concrete ecosystem instantiates one of these networks to explain why the ecosystem exhibits a stable behaviour. The explanation in these cases would work as follows:

> Ecosystem *E* instantiates a network *N* which, in virtue of being of type *X* has the topological property *P*. Therefore, *E* also has *P* (adapted from Huneman 2010).

Interestingly, these types of explanations: (1) do not mention either the entities or the activities that might be going on in the ecosystem, insofar as network analysis only represents relations in terms of the number of nodes and the strength of their interactions—thus being applicable to multiple kinds of systems, just replacing "node" for the objects that are studied in the field (Internet, metabolic networks, social networks, etc.); (2) do not elaborate any kind of causal story that is responsible of producing the phenomenon under investigation. The *explanandum* (*E* having *P*) is accounted for simply because the network is of type *X*, and thus necessarily must instantiate P. Thus, topological properties alone would explain *P* obtaining, and there is no role left for mechanisms (Huneman 2010).

The study of behaviour explanations provided in terms of networks analysis has been centred in the study of explanatory patterns in systems that instantiate non-random networks (e.g. scale-free networks, or small worlds). However, no attention has been devoted to study the explanatory patterns that underlie behaviour explana-

[3] The exact definition of stability is an agitated topic in ecology, and different diversity-stability hypotheses are formulated accordingly (McCann 2000: p. 230, Table 1; Nikisianis and Stamou 2016: pp. 35–36; Gonze et al. 2018: p. 42, Box 1). In most cases, though, a system is qualified as stable if and only if it is able to return to its initial state after a perturbation (resilience), or also the capacity of a population to resist invasions by external species. We will specify later what "stability" means in our case study.

tions for systems that instantiate purely random networks. We suspect that because random networks neither have a particular topology associated, nor an exclusive type of network motifs, the type of explanatory patterns that underlie behaviour explanations of the systems that instantiate them might be slightly different from the types of explanatory patterns provided in terms of non-random networks. The rest of the paper is thus devoted to analyse the explanatory features of a behaviour explanation given for a system that instantiates a random network.

3 Case study: the stability of the human microbiome

The human gut is an ecosystem consisting in a large community of microbes (≈ 1000 species), whose stability behaviour is crucial to maintain human's health.[4] Recent empirical research suggests that the human microbiome exhibits a stable behaviour: even if different individuals might bear different microorganisms in their microbiome, the species that compose the microbiome of an individual, and their relative densities, tend to remain largely stable during her lifetime (Dethlefsen and Relman 2011; Faith et al. 2013). The reasons that make such essential community to behave stably despite the existence of constant perturbations are yet unknown, though. One possible way to explain why the microbiome behaves stably would be to argue that it does so as a consequence of the great number of species that compose it. The explanation in this case would work as follows: insofar as the microbiome is an ecosystem which is composed by a great number of species, and ecosystem biodiversity is believed to foster ecological stability under certain circumstances, then it will be expected that the human microbiome exhibits a stable behaviour (McCann 2000; Ives and Carpenter 2007). This way of accounting for the *explanandum* poses a serious challenge, though: the positive correlation between diversity and stability only works for non-random ecological communities; however, the opposite has been demonstrated to be true for random communities, in which an increase in biodiversity fosters *in*stability (May 1972). Because the microbiome is a random ecological community that due to its biological nature is expected to be suffering constant perturbations, then it will tend to be unstable. Therefore, what ecological theory predicts (instability) and what is empirically observed (stability) are at odds. The question that arises is then the following: what type of dynamics are instantiated in the microbiome so that its interactions result in a stable behaviour?

Fairly recently, Mougi and Kondoh (2012) have elaborated a model that overcomes the difficulty that May's results pose to explain the stability behaviour of random communities. In their view, the problem with May's model is that he only analysed communities with one interaction behavioural type (i.e. where all the members were

[4] In ecology, the concept of "stability" can be used to mean both that the number of species of the microbiome remains constant (i.e. that no species gets extinguished, also called *persistence*), and that the species density in the community recovers quickly after the community has been perturbed (i.e. once the density of one of the species in the community has slightly changed, also called *resilience*). A community whose species density remains constant is said to be in equilibrium. Obviously, if a community is stable in the second sense, it will also be stable in the first sense, but the opposite is not necessarily the case. In the case study that we present here, "stability" refers to the ability of the microbiome to recover its initial species density after a perturbation, i.e. it is a model to study resilience.

either mutualistic, or antagonistic, etc.), and they thought that different results might be obtained if the communities were studied taking into account the fact that there might be different behavioural types interacting simultaneously. They observed that, in fact, the existence of different combinations of interaction types in a community might be a solution to May's results, and thus allows showing that an increase in biodiversity *alone* (i.e. irrespectively of the interacting types) does not trigger instability. Mougi and Kondoh applied their reasoning to macroscopic communities, showing that in communities with different interaction types, an increase in biodiversity does not necessarily foster instability, if the proportion of cooperative types is high, thus explaining stability in terms of cooperation.

Coyte et. al's explanation of the stability of the microbiome follows the same logic as Mougi and Kondoh's research. They agree with them that a key element to explain the stability behaviour of the microbiome is the appeal to the existence of different interacting types. However, contrary to the claim by Mougi and Kondoh that cooperation fosters stability in macroscopic communities, Coyte et al. argue that it is *competition*, *not cooperation*, what explains the stability behaviour of microscopic communities. To prove their claim Coyte et al. develop a mathematical model and a series of computational simulations of the behaviour of the microbiome, and showed that the same result (i.e. that competition stabilizes and cooperation destabilizes communities) was observed irrespectively of the size of the community. Based on their analysis, they hypothesize that the destabilizing effect of cooperation is due to the strong dependencies among species that it generates, which would lead to the appearance of feedback loops in the community, whose destruction would lead the community to collapse. These feedback loops, on the contrary, would not appear when the species in the microbiome compete, thus making the community behaviour more stable. In this section, we analyse the mathematical model that Coyte et al. use to explain the stability of the microbiome.

Coyte et al. model the microbiome as a network of interconnected species to abstractly study its dynamics by applying a combination of linear stability analysis with computational simulations.[5] Their model describes the dynamics of density change of one species X_i given the interaction with another species X_j, and it is written as a modification of the Holling type 1 functional response:

$$\frac{dX_i}{dt} = X_i\left(r_i - s_i X_i + \sum_{j=1, j\neq i}^{S} a_{ij} X_j\right) \quad i = 1, \ldots, S \tag{1}$$

System of Eq. (1) expresses how the density of species i (X_i) will change over time given its own intrinsic growth rate (r_i) its interaction with other members of the same species, i.e. its self-interaction rate (s_i), which it is assumed to be the same for all species (i.e. $s_i = s$), and its interactions with the members of every other species,

[5] Their research consists in three different mathematical methods. In method 1 (linear stability analysis), they only consider communities that are close to equilibrium, while in methods 2 and 3 (permanence analysis, individual-based model) they investigate the behaviour of communities that are far from their equilibrium. Those two later methods yield the same results as the former (cooperation destabilizes communities). For reasons of space, we only consider method 1 for our analysis of the nature of explanation.

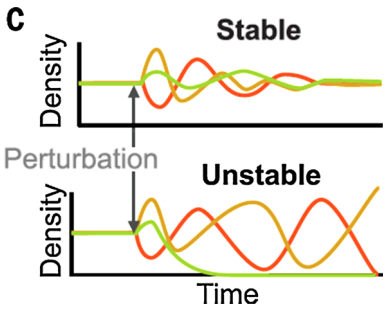

Fig. 2 Species-density variation due to a perturbation. Those communities that return to its equilibrium point are considered stable, and those that return to their previous equilibrium faster after a perturbation are considered more stable. Those that do not return to their equilibrium are deemed unstable. In this image there are two communities represented, stable and unstable. (From Coyte et al. (2015: Supplementary Figure 1C)

or interaction strength (a_{ij}, such that $j \neq i$). Finally, S expresses the total number of interacting species of a given community.

In Coyte et al.'s model, the interaction of a given species within the network will be determined by two parameters: first, the connectivity of i in the network, $C = [0, 1]$, defined as "the fraction of all S species that a single species i interacts with" (Coyte et al. 2015: Supplementary 4). Second, the nature of the interaction types between microbial species. They can take up to five possible forms, based on the signs of a_{ij}/a_{ji}: P_m (cooperation +/+), P_c(competition −/−), P_e (exploitation +/−), P^- (commensalism −/0) and P^+ (amensalism +/0). Accordingly, the proportion of interaction types between species must be established for a given community, being the total proportion of interaction types equal to one:

$$P_m + P_c + P_e + P^- + P^+ = 1 \tag{2}$$

Given this, Coyte et al. elaborate a phase portrait of their model to study its dynamical behaviour. To do so, they start by determining its equilibrium points, i.e. the points for which the variables of the system remain constant over time. Second, they determine the stability of each equilibrium point. An equilibrium point will be stable if after a small perturbation in the system the variables return to the same values that they had before the perturbation, and it will be unstable otherwise (Fig. 2). To determine the stability of each equilibrium point they perform a linear stability analysis. The linear stability analysis for an equilibrium point y is performed in three steps: (i) constructing the Jacobian matrix M of the system. The Jacobian matrix of a N dimensional system is a $N \times N$ square matrix whose elements m_{ij} will be given by the partial derivatives of the system. (ii) evaluating M at the equilibrium point $M|_y$ (iii) computing the eigenvalues of $M|_y$. Once the eigenvalues are computed, the stability of such equilibrium is determined by the following criterion: the equilibrium point y will be stable if and only if the real part of all the N eigenvalues of $M|_y$ is negative.[6]

When working with large dimensional systems the eigenvalues of the Jacobian matrix M tend to follow a concrete distribution. In the case of Coyte et al.'s model,

[6] We will use "stable points" to refer to what mathematically are defined as "asymptotically stable points".

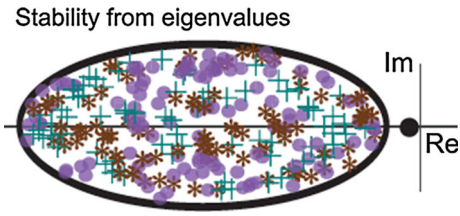

Fig. 3 Distribution of eigenvalues of a microbe community. The largest real part of the eigenvalues determines whether and how quickly the community will return to equilibrium after a perturbation. If this quantity is negative, the community is deemed stable. More negative implies more stability. The imaginary components determine the frequency of the oscillations in population densities after a perturbation. The different colors represent different simulations. (From Coyte et al. 2015: Supplementary Figure 1D)

the eigenvalues that they compute for each equilibrium fall into an ellipse of horizontal radius r_e in the complex plane and centre at $(-s, 0)$, being s the "average self-interaction", except for a single eigenvalue r_s which can lie outside (Fig. 3). Therefore, because an equilibrium requires all its eigenvalues to have a real negative part to be stable, an equilibrium in Coyte et al.'s model will be stable if and only if

$$\max(r_e, r_s) - s < 0 \tag{3}$$

Once we have briefly introduced a basis to study the stability of a given large dynamical system we can focus on analyse our case study. If we denote by y a given equilibria for Coyte's model, the Jacobian Matrix evaluated at that equilibria has the following entries (Coyte et al. 2015: Supplementary 3),

$$m_{ii} = -s \quad i = 1, \ldots, S \tag{4}$$

$$m_{ij} = a_{ij} \quad i, j = 1, \ldots, S$$

so that its eigenvalues—and therefore its stability—depend on the values of the self-interactions s and the a_{ij} terms, i.e. on the type of interactions between the species (cooperation, competition, etc.). Moreover, the connectivity also plays a role in the entries of the Jacobian matrix, so that the lower the connectivity of the network, the more a_{ij} terms will be equal to zero.

Coyte et al. are interested in computing the eigenvalues for each equilibrium point because the ecological stability behaviour of the system is characterized in terms of magnitudes directly related to the eigenvalues distribution of such equilibrium. First, if the equilibrium point is mathematically stable, then it will be ecologically stable too, i.e. the species density before the perturbation took place will eventually be recovered afterwards. In their context, a perturbation is produced when the densities of the species of community changes. Second, the behaviour of the community will be classified as more or less stable depending on how quickly it recovers its initial density distribution after the perturbation. Mathematically, this can be studied by analysing the eigenvalue distribution, such that the more negative the values of the distribution, the more attracting the equilibrium point will be, i.e. the faster the densities will go back to their initial states, and thus the system will be classified as more stable.

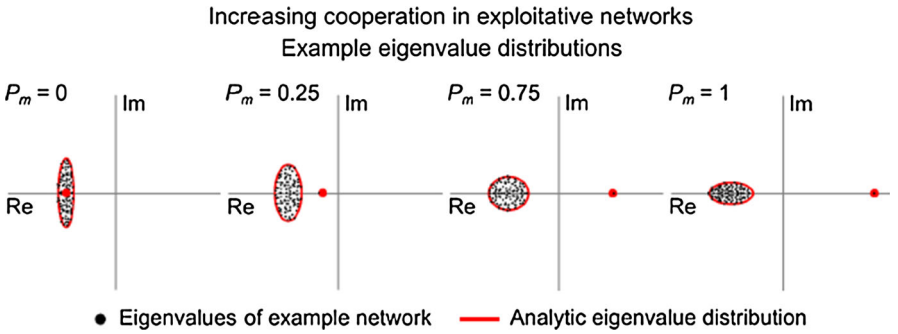

Fig. 4 Numerical confirmation of the analytical model for increasing cooperation in an exploitative community. The image shows how the distribution of the eigenvalues of a given exploitative community changes with respect to the increase in cooperative interactions. According to the linear stability analysis, a system will be more stable if the distribution of the eigenvalues lies further into the negative real part. Here, the single eigenvalue r_s shows how increasing the proportion of cooperative interactions destabilizes the system. (From Coyte et al. 2015: SM. Figure S3)

To study how the degree of ecological stability of each equilibrium point depends on the parameters r_i, s, a_{ij}, in system of Eq. (1), Coyte et al. run a series of simulations with different community types (exploitative, random and competitive). To do so, they define a measure of stability U, such that:

$$U = -\max(r_e - s, r_s) \tag{5}$$

Equation (5) gives the rightmost eigenvalue, i.e. the eigenvalue that represents the lower level of ecological stability. Right after, they study the behaviour of the following derivative $\frac{dU}{dP_m}$, that measures how U varies with respect to the proportion of cooperative interactions P_m. The sign of this derivative describes whether the community behaviour becomes more stable (if it is positive) or less stable (in case it is negative) in function of the proportion of interaction types of the species in the community. With this method, they show that for a given value of C and any given community type, gradually increasing cooperative interactions nearly always increases the overall return rate (the time it takes for the community to recover its initial densities after the perturbation), and the likelihood of the community being unstable (Fig. 4). This method also serves to prove the key role that C plays in the community behaviour: for a constant value of P_m, gradually increasing the value of C always has a destabilizing effect on the community. The overall result of their different simulations is that the higher the values of P_m and the higher the values of C, the less stable the community will be (Fig. 5).

A final step in Coyte et al.'s analysis is the ecological interpretation of their results, i.e. the explanation of what happens in the ecological community when the values of P_m and C increase so as to make the community less stable. In their view, communities with high values of C and P_m might generate strong dependencies among some of its members, which could be transformed into the existence of feedback loops in the community. The existence of feedback loops makes the community sensitive to small perturbations, insofar as a small change in the density of one of the species

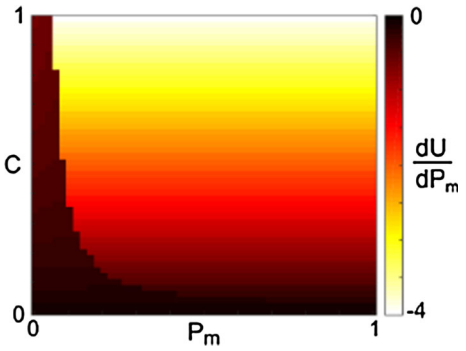

Fig. 5 Effect of increasing cooperation on stability in exploitative communities. This heat map shows how stability varies with respect to connectivity and the proportion of cooperation interactions in an exploitative community. The darker areas show higher stability, whereas the lighters show lower stability. It can be appreciated how stability is higher for the community with almost non-cooperative interactions. High connectivity makes the system unstable when cooperative interactions appear, for the combinations of both makes the system more vulnerable. The combination of low connectivity and high cooperation makes the system stable because of redundancy. (From Coyte et al. 2015: Supplementary Figure S4)

might trigger a cascade effect in the community that will in the end move it towards a different equilibrium point than the one it had before the perturbation. A low value of C and P_m avoids the creation of feedback loops and, therefore, has a stabilizing effect.

Coyte et al. rely on the empirical work done by Stein et al. (2013) to claim that their model is supported by empirical data. After the linear stability analysis and the different simulations they run, Coyte et al. make the following prediction: the proportion of destabilizing cooperative interactions in a stable microbiome has to be low in comparison with competitive and exploitative links, which will be predominant. The data presented by Stein et al. provides empirical validation for such hypothesis, thus suggesting that their model captures the right explanation of the stability of the microbiome.

4 Coyte's explanation as non-mechanist: explaining without providing a causal story

The case study presented above illustrates a type of explanation that we argue does not strictly follow the standard conception of scientific explanation as presented by the new-mechanists. To recapitulate, new-mechanists argue that to explain a phenomenon consist in: first, identifying a model of mechanism (individualized by its entities, its activities and their organization); second, identifying a causal story by means of which the model of mechanism produces the phenomenon to be explained. The question is now to determine in which sense (and to which extent) Coyte et al.'s explanation of the behaviour of the microbiome does not fulfil these two requirements. Particularly, we argue that even if a model of mechanism can be identified in Coyte et al.'s explanation, it is not the case that the explanatory force of the explanation they provide comes from a causal story being told.

Let us start by considering whether Coyte et al. individuate a model of mechanism. As we explained above, Coyte et al. model the microbiome as a network of interacting microbial species that behaves following the dynamics dictated by the system of Eq. (1). In their model, the entities are the different S microbial species that are part of the network, i.e. that compose the microbiome, and whose densities X_i are being studied. The activities of each of the entities are determined by the type of interactions they engage in (competition, cooperation, etc.), and would be given by the sign of a_{ij}. Finally, the organization of the community is given by the random network that describes the interactions among the species, including its number of nodes and its connectivity. Because the elements of the model of mechanism seem to be present in the case of the explanation given by Coyte et al., it seems that the first requirement to have a mechanistic explanation is satisfied by their model.

Once the model of mechanism has been specified it becomes necessary to show how the interactions among the different parts that compose it can produce the *explanandum*. In the case of Coyte et al. the *explanandum* is the stability behaviour of the microbiome, that is, how the species densities remain constant over time despite the existence of perturbations (i.e. some species that vary their densities). The *explanans* says that the key element that makes the microbiome stable is the existence of a high degree of competitive interactions among the species that compose it (so that competition *explains* stability). To get the appropriate connection between the *explanans* and the *explanandum* Coyte et al. proceed as follows: (i) they stipulate a community type (e.g. exploitative community); (ii) they determine its equilibrium points; (iii) they analyse the behaviour of the rightmost eigenvalue in function of the variation of the value of two key parameters of the topology [the connectivity C, and the proportion of interacting behaviours, as defined in (2)] in order to determine the degree of stability of each equilibrium point. In their analysis they observe that, for a fixed value of C, proportionally increasing P_m tends to make the system less stable (ecologically: less resilient). In other words, the time that the system will take to return to its initial equilibrium state after a perturbation will be larger, until a critical value of P_m is reached, such that the return time is equal to ∞, that is, the system becomes unstable.[7]

Now, the question for the new-mechanist is: does Coyte et al.'s mathematical model for the explanation of the stability behaviour of the microbiome actually captures *a causal story* of what happens in the microbiome so that it is ecologically stable? We suspect that the answer to this question is negative. First of all, because the way how Coyte et al. determine its stability behaviour is by means of a linear stability analysis of the system of ordinary differential equations (ODEs) specified in (1). And second, because even once the linear stability analysis is performed, what needs to be modelled to produce the *explanandum* is how the variations in P_m will affect the stability behaviour of the system.

Concerning the first point, the new-mechanist might argue that once the system is modelled, each ODE tells us a different causal story, so that the system as a whole is just an abstraction of the sum of all the individual causal stories modelled by each

[7] In their model, Coyte et al. do not exactly determine at which point the system will become unstable. It is enough for their explanation to show the general tendency of the community to an increasing value of *Pm*.

differential equation. But, even when all these causal stories are put together, they do not lead to the production of the phenomenon. At most, the ODEs give us information about how the variations in the densities of some species will respond to the variations in the densities of others. They do not give any information about how these variations are produced, nor how the variation of the densities of one species will affect the variation of the density of the rest of the species that compose the microbiome. By themselves, they say nothing about how the dependencies among the entities that compose the microbiome produce its stability. The only way in which the ODEs might say something about it is by numerically computing its evolution with respect to time (i.e. the trajectories of the system) when a minimal perturbation occurs.

However, for their explanation, Coyte et al. do not even consider the particular trajectories of the system. Their explanation only requires to study the stability of the equilibria by performing the mathematical steps described before (creating the Jacobian matrix, evaluating it, studying its eigenvalues, etc.), a procedure which does not require to specify the intermediate values of any of the variables X_i of the system. It is enough to study its long term behaviour, no matter which intermediate processes generate it. For this reason, it is difficult (if not impossible) to see how any of the steps followed by Coyte et al. describe a causal story that relates the specific interactions of the entity with the phenomenon to be produced.[8]

In relation to our second point, it seems to us that the explanation by Coyte et al. includes an additional step that makes it even harder to see how the causes can produce the stability behaviour. Concretely, once the ODEs generated by the system of Eq. (1) are solved, Coyte et al. still need to run simulations to determine how the variations in the proportions of the interacting species will affect the stability of the community. This step is crucial, because their *explanans* is precisely that a community will be stable if and only if it has the right proportion of interacting types. This step, as we said, is performed by analysing the variations of the rightmost eigenvalue to an increase in the proportion of P_m. But it is not specified which of the concrete species interactions will become cooperative, nor is that necessary to highlight the negative impact of cooperation on stability. The knowledge of the proportion of species that interact cooperatively is enough to establish their claim. Thus, again, the rehearsal of a causal story seems unnecessary to produce the phenomenon that Coyte et al. are explaining and therefore their explanation is not mechanistic.

5 Explaining with mathematics: combining topology with interaction types to explain stability behaviour

Coyte et al.'s model explains the stability behaviour of the microbiome but it does so in non-mechanistic terms. The question now is to determine how their model gains its explanatory force. In this section, we argue that what makes Coyte et al.'s model explanatory is the combination of the topological properties of the network instanti-

[8] Our argument in this section is inspired by a similar argument presented in Issad and Malaterre (2015: p. 284).

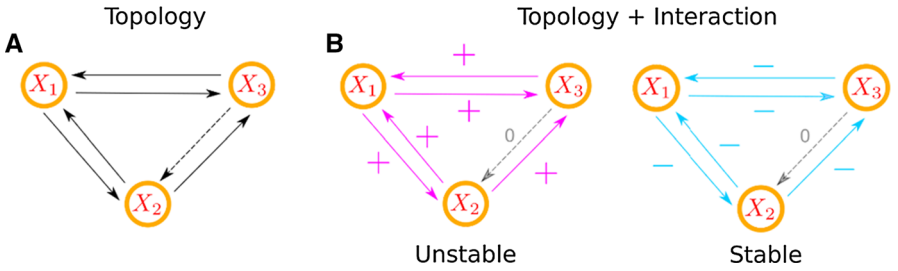

Fig. 6 Schematic representation of the microbiome network. Figure A represents the topology of the network by exclusively pinpointing the interactions among the species (the dashed line between X_3 and X_2 indicating that $a_{32} = 0$). Figure B represents the effect that considering not only the topology, but also the interaction types, might have on stability: for two communities with the same topology, only the one where competition is high (right) would be stable

ated by the microbiome with the knowledge about the dynamics that each of these topologies acquires in virtue of the combination of interacting types in the community.

Let us start by discussing the role of the topology in their *explanans*. In Coyte et al.'s model, the knowledge of the parameters s_i and a_{ij} allows determining the value of C and thus provides the topology of the system. Notice that if our network were non-random, i.e. if it realized a concrete topology, at this point we would have all the necessary information to decide whether the network is stable or unstable. Take the example of a scale-free network: were the microbiome a scale-free network, we could already identify its hubs (these nodes whose alteration would destabilize the network), as well as how it will behave if we increase the number of nodes, if we destroy some of the edges between nodes, etc. If this were the case, Coyte et al.'s explanation would be much simpler: the microbiome behaves stably *because* it realizes a scale-free network. This type of explanation would have the same nature as other topological explanations, gaining its explanatory force simply from the nature of the network that the system instantiates (Huneman 2010, 2018b; Jones 2014).

However, contrary to what happens in non-random networks, the microbiome instantiates a random network, which means that none of its topological properties will be known a priori. For that reason, Coyte et al. need to study the dynamics that the topology instantiates in order to determine which are the conditions that will make it stable. Furthermore, in their case, what ultimately explains whether the dynamics instantiated will be stable is the proportion of interacting types in the microbiome. That is, not every community with the same topology will be equally stable. Their explanation, therefore, needs to combine the determination of the topology of the system with the study of the dynamics that the variables that instantiate that topology will have under different conditions, i.e. for different interaction types (Fig. 6).

Therefore, Coyte et al.'s explanation of the stability behaviour of the microbiome would consists in something like:

> Ecosystem E instantiates a network N such that: (i) N corresponds to one of a set of topologies φ, and (ii) φ has the right proportion of interaction types. Thus, N has property P. Therefore, E also has P

Let us now develop what this model of explanation consists in. First, concerning (i), it is possible to mathematically compute which networks will make the microbiome stable, but this knowledge alone will render the *explanandum* unexplained, because there is still an element missing: which is the property in virtue of which N's topology is stable? Or, in other words, why N has a dynamics such that the system will tend to keep its stability? Appealing to its topology, by itself, will undermine the answer, because the system is purely random and there are multiple states that will make it stable, as well as multiple states that will make it unstable. But this option is not possible in the case of purely random networks, and therefore the explanatory force of Coyte et al.'s model, even if requiring the appeal to the topological structure of the microbiome, needs to be acquired from somewhere else.

Second, about (ii), our proposal is that the key additional element that makes Coyte et al.'s model fully explanatory of the stability behaviour of the microbiome is their appeal to the different interaction types that might appear in the network. The reason is that for the microbiome to be stable it is necessary that the *dynamics* instantiated by the topology are conducive to a stable equilibrium state. The only way of showing what are the conditions under which the dynamics instantiated by a random topology are conducive to a stable state is by studying their response to slight modifications in the parameters that define the system, namely a_{ij}. The way of computationally studying this is by evaluating the response(s) of the rightmost eigenvalue to changes in the values of a_{ij} in relation to changes in the proportions of interacting types in the community. In other words, it is necessary to study how the changes in the proportions of cooperative, competitive, exploitative, etc. interactions will affect the stability of the microbiome. Only once this response is studied, and the results are analysed, will the model be fully explanatory of the stability behaviour of the microbiome. In this sense, even if knowing the topology is necessary, since the dynamics of the system directly depends on the topology it instantiates, its knowledge is not sufficient to explain its behaviour. It is necessary to additionally understand how the community will respond to different proportions of interaction types. Only after this is done will the model gain its explanatory force, and Coyte et al. can assert that competition *explains* the stability behaviour of the microbiome.

Even if the points we have made here are relative to the case study we have analysed, we suspect that our conclusions about the nature of explanation can be extended to all the cases of behaviour explanations when they are given in terms of random networks. One key feature that distinguishes random from non-random networks is that the latter, but not the former, have specific properties derived from their network motifs and that can be unequivocally ascribed to every system that instantiates them. This does not mean however that random networks do not have network motifs: it means that their network motifs will highly depend on the way in which the network is constructed, and it will not be constant among all the systems that instantiate a random network. Because of this, we suspect that explanations of the behaviour of systems that instantiate a random network can only be given by studying an alternative mathematical property of the network (e.g. in Coyte et al.'s, the dynamics of the network), because: first, their network motifs will only be known once its mathematical properties have been studied; second, they will be highly variable depending on the mathematical properties that the network instantiates. This sharp contrast between random and non-random

networks is thus fundamental to understand why explanations in terms of the latter can be exclusively topological, whereas in the case of the former the topology has to be complemented with an appeal to dynamics.

6 A case for integrative pluralism in behaviour explanation

Until now we have argued that Coyte et al.'s explanation of the stability behaviour of the microbiome is neither purely mechanistic, since it lacks a causal story, nor exclusively topological, since the topology needs to be combined with a dynamics to be fully explanatory. We now argue that our case study supports an integrative pluralistic picture of behaviour explanation in biology. Following Mitchell (2003) and Brigandt (2010, 2013b; Brigandt et al. 2017), we take an explanation to be integrative when it requires the combination of concepts from different fields and of different types in order to gain its explanatory force. Such integration is normally driven by pragmatic considerations about the question asked: that is, as some scientific questions are about very complex phenomena, their responses usually require the integration of the knowledge of different fields to be fully satisfactory (Brigandt 2013a). In this sense, integrative pluralism is both beyond simple pluralism and against reductionism. Like simple pluralism, integrative pluralism accepts that some phenomena in ecology might be explained by simply telling a causal story (i.e. describing a mechanism), whereas others might be exclusively given in mathematical terms. However, integrative pluralism tries to go beyond this simple idea by showing how some complex phenomena are explained by integrating knowledge from different fields (ecology, population genetics, molecular biology, etc.), using both, causal mechanistic strategies and mathematical modelling. Against reductionism, integrative pluralism embraces the idea that even if the explanation of a complex phenomenon requires the appeal to knowledge from different fields, all of them are indispensable for the *explanans* to account for the *explanandum*. In other words, that the explanation of the phenomenon is only possible by integrating all the knowledge provided by the different fields, which would not be obtained if the knowledge of one of the fields were reduced to the knowledge of some of the others.

Following the tenets of integrative pluralism, we will make the point that the explanatory model of the stability behaviour of the microbiome presented by Coyte et al. exemplifies an integrative explanation. Concretely, their explanation combines a model of a mechanism with a mathematical model plus a series of computational simulations, integrating mathematical knowledge (network modelling, linear stability analysis), with knowledge about the patterns of interactions in ecological communities. We further argue that the necessity to integrate knowledge about the model of the mechanism with knowledge about the mathematical properties of the community (including its topology) is common for every scientific explanation that accounts for the phenomenon in terms of random networks.

As we argued in Sect. 4, Coyte et al.'s *explanans*, despite not telling any causal story of what happens in the microbiome, describes a model of the mechanisms enumerating which are the interacting entities, what type of activities they engage in, and the type of organization that the microbiome has. Also, as we argued in Sect. 5, their explanatory

model relies on the dynamics of the microbiome, which is acquired in virtue of the topology that it instantiates. The study of the behaviour of this dynamics determines a set of possible topologies, any of which will be stable, and thus one of the possible states the microbiome could be at. Importantly, both the model of the mechanism and the mathematical model are necessary and none of them is by itself sufficient without the other for Coyte et al.'s *explanans* to account for the *explanandum* in terms of "competition", as the authors say it does.

First of all, the model mechanism without the dynamics is completely unspecific about the stabilizing role of competition in the microbiome. Since they are explaining a behaviour (stability behaviour), and the behaviour is the result of the set of interactions between the entities (the species that compose the microbiome), it is necessary to study the way in which the interactions result in the behaviour that is observed. Or, in other words, detailing the entities that interact and the activities they engage in is insufficient to explain the stability behaviour if the way how these interactions make the system change is not studied adequately. For sure, the model of the mechanisms specifies that competition is one of the activities that the entities of the microbiome engage in and that can affect its stability. However, because the model of the mechanism in itself does not describe how the system changes in time, its specification is insufficient to make competition explanatory.

Secondly, the topology of the system, without the knowledge of the model of the mechanism, is also insufficient to explain why the microbiome exhibits a stable behaviour. As the community instantiates a *random* network, its network motifs have to be studied by analysing its dynamics. In the case of Coyte et al. they perform a linear stability analysis to study the reaction of the community to small perturbations. This analysis, however, does not make competition explanatory of the stability behaviour of the microbiome. As we explained extensively in Sect. 5, once Coyte et al. have determined the distribution of a_{ij} that make the microbiome stable, they have to study how the variations in the proportions of interaction types will affect the stability of the community. In other words, what makes competition explanatory in Coyte et al.'s model is not the strength of the interactions between the members of the microbiome, but the character (competitive, cooperative, exploitative, etc.) that those interactions have.

The previous observation entails that it is possible to have two communities with the same topology (i.e. with the same nodes, edges), but where only one of them is stable, whereas the other is not, due to the role that the interaction types have on the stability behaviour of the microbiome. Importantly, the knowledge of the influence of the interaction types on the dynamics that the topology instantiates is only possible once the model of the mechanism has been specified. Before this happens, and cooperation, competition, exploitation, etc. have been defined as possible activities of the entities, it is impossible to know whether the interaction types will have any influence on the dynamics of the network, since the latter is not the case for many other networks that might instantiate a random topology (e.g. the internet). Therefore, the knowledge of the dynamics is, by itself, insufficient to make competition explanatory, since the knowledge that competition will influence stability derives from the knowledge of the model of the mechanism of the microbiome.

From the two previous observations we derive that the explanation of the stability behaviour of the microbiome is a case of integrative pluralism, where a model of the mechanism needs to be combined with a mathematical analysis in order to be explanatory. As we argued in Sect. 5, the stability behaviour of the microbiome cannot be explained exclusively in terms of the topological properties of the network. Since it is a random network, it needs to be complemented with the study of the dynamics that the topology instantiates. As we argued there, this is a consequence of the fact that the network motifs of a random network can only be discovered a posteriori, by determining the effects that the interaction types have on the stability of the network. However, this claim about the study of the dynamics cannot be extended to every random network, since the reasons that make each random network stable will be different depending on their nature, and thus on the type of mathematical analysis that has to be done to determine its properties. The type of mathematical analysis will nevertheless depend on the model of the mechanism that is instantiated in each case, which will determine the nature of the entities that interact, as well as how their activities will be produced. The point we are making is thus that every scientific explanation that relies on the realization of a topology will be a case of integrative pluralism if the network that is realized is random.

7 Conclusion

The case study analysed in this paper fits well with the recent tendency in philosophy of science to emphasize the important role that mathematics play in some scientific explanations. Particularly, our case study, even if focused on behaviour explanations, shares many similarities with some of the cases of topological explanation analysed by Huneman (2010, 2018a, b). As in the cases he studies, the stability of the microbiome cannot be explained purely in mechanistic terms, due to the impossibility of telling a causal story that explains how the system behaves. Additionally, the explanation we analyse in this paper also gains its explanatory force from the specification of the topological properties of the system. However, and in contrast with the cases of explanations of the behaviour of non-random networks, in the case presented by Coyte et al. (which analyses the behaviour of a random network) the specification of the topology of the community is not enough to account for the *explanandum* (the stability behaviour of the microbiome). The authors are also required to study the dynamics that the topology instantiates, as well as to discover which set of topologies will make the system stable in virtue of its dynamics. Such set of topologies can only be discovered by studying the response of the rightmost eigenvalue to different proportions of P_m. Because the knowledge that the interaction types will influence the stability behaviour of the microbiome is only acquired after the model of the mechanism for the phenomenon is expelled out, Coyte et al.'s explanation constitutes a case of integrative pluralism. In other words, the explanatory force of their model is only gained from the combination of mathematical and mechanistic knowledge.

As a consequence, the analysis of our case study proves two main points: (a) the explanation of the behaviour of non-random ecological networks is different form the explanation of the behaviour of random networks. That is to say, because the network

is random, there is not any network type whose specification would automatically explain the behaviour of the system, and thus the network motifs have to be discovered; (b) behaviour explanations of random ecological networks are cases of integrative pluralism, in which knowledge from mathematics and from ecology are integrated to solve a complex problem. Concretely, the model of the mechanisms determines how to construct and study the mathematical model so that the *explanans* that accounts for the *explanandum* can be provided.

Acknowledgements A previous version of this paper was presented at the IX Conference of the Spanish Society for Logic, Methodology and Philosophy of Science (UNED, Madrid, 2018). We acknowledge all the participants for their helpful suggestions. José Díez, Johannes Findl, Christophe Malaterre, Eric Muszynski, and two anonymous reviewers provided insightful comments on a previous version of the paper. APC and JS are especially thankful to Sabina Leonelli, who had the idea of bringing together researchers from the Living Systems Institute and Egenis (University of Exeter), and thanks to whom the authors met and started to think about the issues that gave origin to this paper. The following institutions are formally acknowledged: RD and JS, Spanish Ministry of Economy and Competitiveness (FFI2016-76799-P); RD, Spanish Ministry of Economy and Competitiveness (BES-2013-063239); JS, Spanish Ministry of Education (FFU16/02570). APC, Spanish Ministry of Economy and Competitiveness (MTM2012-31714).

References

Alleva, K., Díez, J. A., & Federico, L. (2017). Models, theory structure and mechanisms in biochemistry: The case of allosterism. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 63,* 1–14.

Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind, 114,* 223–238.

Baker, A. (2009). Mathematical explanation in science. *The British Journal for the Philosophy of Science, 60,* 611–633.

Baker, A. (2015). Mathematical explanation in biology. In P. A. Braillard & C. Malaterre (Eds.), *Explanation in biology: An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 229–247). Dordretcht: Springer.

Bechtel, W. (2010). The downs and ups of mechanistic research: Circadian rhythm research as an exemplar. *Erkenntnis, 73,* 313–328.

Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science, 78*(4), 533–557.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 421–441.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as scientific research strategies*. Cambridge, MA: The MIT Press.

Brigandt, I. (2010). Beyond reduction and pluralism: Toward an epistemology of explanatory integration in biology. *Erkenntnis, 73*(3), 295–311.

Brigandt, I. (2013a). Explanation in biology: Reduction, pluralism, and explanatory aims. *Science and Education, 22*(1), 69–91.

Brigandt, J. (2013b). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 44,* 477–492.

Brigandt, J. (2015). Evolutionary developmental biology and the limits of philosophical accounts of mechanistic explanation. In P. A. Braillard & C. Malaterre (Eds.), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences* (pp. 135–173). Dordretcht: Springer.

Brigandt, J., Green, S., & O'Malley, M. A. (2017). Systems biology and mechanistic explanation. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy (Chapter 27)*. London: Routledge.

Coyte, K. Z., Schluter, J., & Foster, K. R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Science, 350*(6261), 663–666.

Craver, C. F. (2007). *Explaining the brain*. New York: Clarendon Press.
Craver, C. F., & Darden, L. (2013). *In search for mechanisms: Discovery across the life sciences*. Chicago: University of Chicago Press.
Dethlefsen, L., & Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *PNAS USA, 108*(1), 4454–4461.
Deulofeu, R., & Suárez, J. (2018). When mechanisms are not enough. The origin of eukaryotes and scientific explanation. In A. Christian, D. Hommen, N. Retzlaff, & G. Schurz (Eds.), *Philosophy of Science. European Studies in Philosophy of Science* (Vol. 9). Cham: Springer.
Díez, J. A. (2014). Scientific w-explanation as ampliative, specialised embedding: A neo-hempelian account. *Erkenntnis, 79,* 1413–1443.
Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., et al. (2013). The long-term stability of the human gut microbiota. *Science, 341*(6141), 1237439.
Foster, K. R., Schluter, J., Coyte, K. Z., & Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem of a leash. *Nature, 548,* 43–51.
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(S3), S342–S353.
Gonze, D., Coyte, K. Z., Lahti, L., & Faust, K. (2018). Microbial communities as dynamical systems. *Current Opinion in Microbiology, 44,* 41–49.
Human Microbiome Project Consortium. (2012). Structure, function and diversity of the health human microbiome. *Nature, 486*, 207–214.
Huneman, P. (2010). Topological explanations and robustness in biological sciences. *Synthese, 177,* 213–245.
Huneman, P. (2018a). Outlines of a theory of structural explanation. *Philosophical Studies, 175*(3), 665–702.
Huneman, P. (2018b). Diversifying the picture of explanations in biological sciences: Ways of combining topology with mechanisms. *Synthese, 195,* 115–146.
Huneman, P. (2018c). Realizability and the varieties of explanation. *Studies in History and Philosophy of Science*. https://doi.org/10.1016/j.shpsa.2018.01.004.
Issad, T., & Malaterre, C. (2015). Are dynamic mechanistic explanations still mechanistic? In P. A. Braillard & C. Malaterre (Eds.), *Explanation in biology: An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 265–292). Dordretcht: Springer.
Ives, A. R., & Carpenter, S. R. (2007). Stability and diversity of ecosystems. *Science, 317*(5834), 58–62.
Jones, N. (2014). Bowtie structures, pathway diagrams, and topological explanations. *Erkenntnis, 79*(5), 1135–1155.
Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science, 78*(4), 601–627.
Kuorikoski, J. (2007). Explaining with equilibria. In *Rethinking explanation* (pp. 149–162). Dordrecht: Springer.
Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science, 64,* 485–511.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Marchesi, J. R., & Ravel, J. (2015). The vocabulary of the microbiome research: A proposal. *Microbiome, 3,* 31.
May, M. (1972). Will a large complex system be stable? *Nature, 238,* 413–414.
McCann, K. S. (2000). The diversity stability debate. *Nature, 405,* 228–233.
Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.
Montoya, J. M., & Solé, R. V. (2002). Small world patterns in food webs. *Journal of Theoretical Biology, 214*(3), 405–412.
Mougi, A., & Kondoh, M. (2012). Diversity of interaction types and ecological community stability. *Science, 337*(6092), 349–351.
Nicholson, D. J. (2012). The concept of mechanism in biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43*(1), 152–163.
Nikisianis, N., & Stamou, G. P. (2016). Harmony as ideology: Questioning the diversity-stability hypothesis. *Acta Biotheoretica, 64*(1), 33–64.
Potochnik, A. (2015). Causal patterns and adequate explanations. *Philosophical Studies, 172*(5), 1163–1182.
Rathkopf, C. (2018). Network representation and complex systems. *Synthese, 195,* 55–78.

Rice, C. (2012). Optimality explanations: A plea for an alternative approach. *Biology and Philosophy, 27*(5), 685–703.
Rice, C. (2015). Moving beyond causes: Optimality models and scientific explanation. *Noûs, 49*(3), 589–615.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
Sober, E. (1983). Equilibrium explanation. *Philosophical Studies, 43*(2), 201–210.
Solé, R. V., & Montoya, J. M. (2001). Complexity and fragility in ecological networks. *Proceedings of the Royal Society B, 268*(1480), 2039–2045.
Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences, 8*(9), 418–425.
Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., et al. (2013). Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. *PLoS Computational Biology, 9,* e1003388.
Suárez, J., & Deulofeu, R. (unpublished manuscript). Equilibrium explanation as structural non-mechanistic explanations: The case of long-term bacterial persistence in human hosts.
Walsh, D. A. (2015). Variance, invariance and statistical explanation. *Erkenntnis, 80*(3), 469–489.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.
Woodward, J. (2017). Scientific explanation. In E. N. Zalta (Ed.), *The stanford encyclopaedia of philosophy*. https://plato.stanford.edu/archives/fall2017/entries/scientific-explanation/. Accessed 1 Nov 2018.
Xia, T. (2010). Network modelling in systems biology. Ph.D. Dissertation. https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=2611&context=etd. Accessed 16 Oct 2018.

Affiliations

Roger Deulofeu[1] · Javier Suárez[1,2] · Alberto Pérez-Cervera[3]

[1] Logos – Research Group in Analytic Philosophy, University of Barcelona, Barcelona, Spain

[2] Egenis – The Centre for the Study of Life Sciences, University of Exeter, Exeter, UK

[3] Department of Mathematics, Universitat Politècnica de Catalunya, Barcelona, Spain