





QUADERNS  
*de Filosofia*

VOL. VI : I  
VALÈNCIA, 2019

*SOCIETAT DE FILOSOFIA DEL PAÍS VALENCIÀ*

## QUADERNS DE FILOSOFIA

*Director:* Tobies Grimaltos (Universitat de València)

*Secretari de redacció:* Sergi Rosell (Universitat de València)

*Consell de redacció:* Ramón Feenstra (Universitat Jaume I), Elsa González Esteban (Universitat Jaume I), Javier Gracia (Universitat de València), Eduardo Rivera López (Universidad Torcuato di Tella), Pablo Rychter (Universitat de València), Jennifer Saul (University of Sheffield), Faustino Oncina (Universitat de València), José Zalabardo (University College London)

*Consell editorial:* Jesús Alcolea (Universitat de València), Vicent Baggetto (IES Joan Fuster de Sueca), Tyler Burge (UCLA), Enric Casaban (Universitat de València), Pascual Casany (València), Jesús Conill (Universitat de València), Adela Cortina (Universitat de València), Christine Chwaszcza (Universität zu Köln), Antoni Defez (Universitat de Girona), Román de la Calle (Universitat de València), Vicente Domingo García Marzá (Universitat Jaume I), Manuel García-Carpintero (Universitat de Barcelona), Tomás Gil (Technische Universität Berlin), Susan Haack (University of Miami), Christopher Hookway (University of Sheffield), Joan Llinares (Universitat de València), Genoveva Martí (Universitat de Barcelona, ICREA), Carlos Pereda (UNAM), Fernando Miguel Pérez Herranz (Universitat d'Alacant), Giovanna Pinna (Università del Molise), Xavier Serra (IES Joan Fuster de Sueca), Sergio Sevilla (Universitat de València), Ernest Sosa (Rutgers University), Nicholas White (UC Irvine)

Edita: Societat de Filosofia del País Valencià

eISSN: 2341-3042

*Disseny i seguiment editorial:* Antoni Domènech

*Col·laborador en la correcció lingüística:* Guillem Llop

© de l'edició, Societat de Filosofia del País Valencià

© dels articles, els seus autors

<http://ojs.uv.es/index.php/qfilosofia>

Redacció: Societat de Filosofia del País Valencià

Universitat de València — Facultat de Filosofia i Ciències de l'Educació

Departament de Filosofia

Av. Blasco Ibáñez, 30-5a. 46010 València

e-mail: [sfpv@uv.es](mailto:sfpv@uv.es)

<http://www.uv.es/sfpv/>

QUADERNS DE FILOSOFIA és una revista acadèmica dirigida als professionals de la filosofia de periodicitat semestral. Publica articles de temàtica filosòfica de qualitat científica i té un interès especial en la didàctica de la filosofia.

Cada número ordinari té tres parts: 1. Articles d'investigació: contribucions originals pertanyents als diversos àmbits de la recerca filosòfica; 2. L'anomenada "Brúixola filosòfica", feta d'articles que presenten un determinat tema o debat filosòfic; i 3. Documents, entrevistes, notes de discussió i ressenyes.

Tots els articles originals són sotmesos a una revisió cega anònima per experts en l'àrea corresponent.

La revista es publica en versió electrònica d'accés obert (Open Journal System).

La revista és inclosa en alguns dels índexs de major prestigi internacional en filosofia i humanitats, com són ERIH PLUS, Latindex, ISOC, Dialnet, Philosopher's Index, DOAJ, EBSCO, REDIB, i el consell de redacció treballa per a la inclusió en d'altres, en especial SCOPUS, FECYT i CARHUS Plus+.

### *Normes de publicació*

Els articles poden estar escrits en català, castellà i anglès, i no han d'excedir de les 10.000 paraules. Han de dur títol i incloure un resum (de fins a 150 paraules) i unes paraules clau en la llengua de l'article i en anglès. Els manuscrits s'han de presentar de forma anònima per tal de garantir-ne el dictamen cec; és responsabilitat de l'autor eliminar qualsevol autoreferència.

També es poden proposar ressenyes de llibres publicats en els darrers tres anys, amb una extensió màxima de 2.000 paraules, llevat de casos excepcionals prèviament acordats amb la direcció de la revista. Les ressenyes no tindran títol ni notes. Com a encapçalament cal consignar les dades completes del llibre ressenyat, inclòs l'ISBN. Cal recordar que una ressenya no és un resum, sinó un examen crític del llibre ressenyat.

Les normes d'estil poden consultar-se en la versió electrònica de la revista, però en tot cas sols s'exigirà el seu compliment en la versió final de l'article o ressenya, si se n'accepta la proposta.



# ÍNDIX

## *Articles. Monogràfic filosofia de la biologia*

“Introduction”, Cristina VILLEGAS & Laura NUÑO DE LA ROSA (UCM) .....	11
“Intervals of quasi-decompositionality and mechanistic explanations”, Emilio CÁCERES (UNED) .....	15
“Indispensability and Effectiveness of Diagrams in Molecular Biology”, Javier ANTA (UB) .....	29
“How not to Resist the Natural Kind Talk in Biology”, María José FERREIRA (UBA) .....	47
“A metaphysical approach to holobiont individuality: Holobionts as emergent individuals”, Javier SUÁREZ (UB) & Vanessa TRIVIÑO (URJC) .....	59

## *Brúixola filosòfica*

“Quines normes? Una aproximació al debat sobre el concepte de malaltia mental”, Virginia BALLESTEROS (UV) .....	79
---	----

## *Ressenyes*

Julien CANAVERA, <i>Pensar problemàticament. Ensayo sobre Gilles Deleuze</i> , per Raimon RIBERA AÑÓ (UV) .....	121
---	-----





# *Articles*



CRISTINA VILLEGAS & LAURA NUÑO DE LA ROSA  
*MSR Research Group, Universidad Complutense de Madrid*

## Introduction

THE RESEARCH WORKSHOP ON PHILOSOPHY OF BIOLOGY AND COGNITIVE SCIENCES (PBCS) has been running annually for eight years now. Since its origins in 2011, the PBCS aims at bringing together young researchers from different disciplinary backgrounds: philosophers, cognitive scientists, and biologists working on issues of common interest. As a result, it has become a reference event in Spain for PhD and early postdoc scholars interested in the philosophy of the life and cognitive sciences. The PBCS enhances research through the discussion of each session in a friendly and interdisciplinary atmosphere intended to debate work in progress. In this issue, we offer a selection of the papers presented at the eighth edition of the event, which took place in May 2018 at the Complutense University of Madrid.

The topics in this issue range from general epistemic considerations to metaphysical concerns in the philosophy of biology. The four selected articles are a good instantiation of how classical topics of interest among philosophers of biology connect with new philosophical frameworks, developed in the light of new empirical evidence and the consequent proliferation of biological research fields. The first two papers concern two distinct epistemological issues, namely mechanistic explanations and diagrammatic representations, in the life sciences. From their different perspectives, both of them address the role of epistemic purposes in structuring scientific explanations in the biological practices. The following two papers engage with two classical metaphysical questions in the philosophy of biology, namely the existence and nature of biological natural kinds and the ontological status of biological individuals, in view of recent philosophical theories accounting for new kinds of biological entities such as developmental homologues and holobionts.

The article by Emilio Cáceres, entitled ‘Intervals of quasi-decompositionality and mechanistic explanations’, deals with recent mechanistic approaches to scientific explanation, which have been especially forceful in the

philosophy of biology literature. Cáceres claims that although it is usual to characterize science as a process of describing mechanisms, there is no consensus among philosophers over what a mechanism is. The purpose of his paper is to analyse certain problems arising from classical approaches, and to develop a novel, pragmatistic account of mechanisms. After distinguishing the main aspects common to most mechanistic views (namely, composition by entities and activities, causal role, and organization), the author encounters a difficulty met by all of them: the definition of a mechanism implies referring to a specific level of organisation whose relation with other levels may be problematic. According to Cáceres, the attempt to classify levels of organisation of nature in an exhaustive way through a hierarchy of mechanisms entails a variety of problematic issues, such as the emergence of higher-level properties or the problem of downward causation. He thus proposes that the particular sciences establish different criteria of *quasi-decompositionality* of systems according to their epistemic goals. This is particularly interesting for biological sciences, where each field defines its objects of study according to the different levels of biological organization. The study of photosynthesis in ecology, plant physiology and molecular biology illustrates this tenet. The result of Cáceres' approach is a view of science as a "nesting of boxes" defined at different levels, where causal interactions only take place at the fundamental one, but non-causal relations between levels can still be established with heuristic purposes.

Javier Anta's article, 'Indispensability and effectiveness of diagrams in molecular biology', concerns a dispute that has recently become a matter of attention among philosophers of biology, namely the epistemic status of diagrammatic representations. In particular, Anta argues that the extended use of diagrams in molecular biology is not merely illustrative in purpose, but plays an important epistemic role. On the one hand, the author defends that diagrams are an indispensable epistemic tool for representing complex properties of molecules. Since they are a well-defined, syntactically-behaved and semantically-driven means of representation, diagrams are able to account for a number of spatial and structural properties that are not always suited for translation into natural languages. Against the thesis that diagrams merely represent in a visual way those properties that would be too hard to account for by linguistic means, Anta argues that some of such properties cannot be computationally processed at all. Therefore, complex structural properties of biomolecules are necessarily coded in diagrammatic schema. On the other hand, Anta also advocates the use of diagrams even in those cases where all the represented properties can be computationally processed. He defends that, in the context of molecular biology, diagrams are more representational and

more inferentially effective than formula-based explanations. He relies on the semantic and epistemic effectiveness of these representations, for they make relevant information more available than by other means. Importantly, the graphical representation of geometric and topological properties makes it possible to encode relevant structural information without inflating syntax exponentially or losing operational rigor. Finally, the article deals with how these epistemic values of diagrams may explain their recurrent and multifacet use in molecular biology: in developing several types of explanations (either functional-structural, mechanistic-dynamic or topological), in testing novel hypothesis, and in predicting new phenomena.

In her paper “How not to resist the natural kind talk in biology”, María J. Ferreira Ruiz argues that two major recent criticisms to the conception of biological natural kinds as property clusters are ill formulated, since they depend on underlying, though unnecessary, assumptions on what a philosophical theory of natural kinds should be able to provide. The first objection concerns explanatory limitation. According to this objection, the cluster approach is flawed because it fails to answer the question of why a natural kind is characterized by certain properties rather than others. The second objection addresses the inability of property clusters views to fix kind membership, and their consequent reliance on independent classificatory criteria. Ferreira argues that these two objections rely on the unjustified assumption that a philosophical theory of natural kinds needs to provide a solution to both requirements. In contrast, she contends that the aim of any theory of natural kinds is not to be able to explain why these kinds are as they are, neither to fix membership. Instead, the only primary requirement one can demand from such a theory is to be able to distinguish natural kinds from non-natural ones. While remaining metaphysically agnostic on the true nature of biological natural kinds, Ferreira forcefully shows how our very expectations from philosophical theories have conditioned this classical debate.

The ontological status of the symbiotic associations between hosts and microbiota has become the source of one of the most lively controversies in the philosophy of biology, polarised between those advocating the individuality of “holobionts”, and those arguing that these associations have a looser nature, analogous to that of ecological communities. In their paper, “A metaphysical approach to holobiont individuality: Holobionts as emergent individuals”, Javier Suárez and Vanessa Triviño align with the former position, but take an original approach that combines new insights from the metaphysical literature on emergence with recent debates in the philosophy of biology on the units of selection. The authors focus on the most controversial hurdle to understand holobionts as units of selection, namely their ability to establish inheritance

relations. They argue that, if inheritance is understood on the basis of trait-recurrence, and given the ability of holobionts to reconstruct their traits intergenerationally by recruiting the members of their microbiota, holobionts can be conceived as emergent individuals. According to the metaphysician Jessica Wilson, higher-level properties of a system are emergent properties in a strong ontological sense if they have new causal powers that are not present in the lower-level properties upon which they depend. The authors apply this criterion to the holobiont case, and conclude that the ability of holobionts to sustain trait recurrence belongs to holobionts as emergent individuals, insofar as some of their traits have a new, downwardly exerted causal power, that determines the relative survival of some of their microbial parts.

The articles of this special issue demonstrate that the proliferation of entities and modes of explanation produced by the life sciences in the last few decades is followed up closely by young philosophers of biology. The heterogeneity of biology is therefore manifested in the philosophical issues here presented, showing that evolutionary theory is no longer the only source of conceptual puzzles in the philosophy of biology, and that new explanatory tools and research objects multiply together with different epistemic purposes.

EMILIO CÁCERES VÁZQUEZ<sup>1</sup>  
*UNED*

# Intervals of quasi-decompositionality and mechanistic explanations

Received: 6/7/2018. Accepted: 17/12/2018

**Abstract:** It is commonly assumed that the concept of mechanism is a keytool for the scientific understanding of observable phenomena. However, there is no single definition of mechanism in the current philosophy of science. In fact, philosophers have developed several characterizations of what seemed to be a clear intuitive concept for scientists.

In this paper, I will analyze these philosophical conceptions of mechanism, highlighting their problematic aspects and proposing a new mechanistic approach based on the idea that the pertinent levels of organization for a mechanistic explanation can be identified with intervals of quasi-decompositionality. I argue that this approach allows us to consider that activities are directly derived from the entities' structure. Consequently, a mechanistic explanation implies an arbitrary but not capricious choice of an organizational level. According to this approach, inter-level causation is merely apparent and there is no place for emergent properties.

**Keywords:** mechanisms, emergent properties, quasi-decompositionality, systems.

## INTRODUCTION

**F**ROM A SCIENTIFIC PERSPECTIVE, observable phenomena are the result of the interactions of the elements underlying a given phenomenon. Thus, the luminous phase of photosynthesis is produced by a series of interactions

<sup>1</sup> Department of Logic, History and Philosophy of Science, UNED, Spain. Mail to epifumi@epifumi.com.

among certain components of the green parts of plants. Explaining how photosynthesis takes place consists of defining, describing and locating the elements involved in that process and the relationships between them. For scientists, this is considered to describe a mechanism and, by extension, *to do science*. (MACHAMER *et al.* 2000, 2)

This claim, which could be summarized by saying that a mechanism *is a set of parts that interact among themselves to give rise to a phenomenon*, does not have, from a philosophical perspective, a single characterization. Instead, there are several different descriptions that attempt to capture the way scientists work. For example, (a) Machamer, Darden and Craver (hereafter MDC) affirm that “*mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions*” (MACHAMER *et al.* 2000, 3); for Glennan (b) A mechanism is a set of parties whose interaction gives rise to a phenomenon, where such interaction “*can be characterized by direct, invariant and change-related generalizations*” (GLENNAN 2002, S334), or its most recent minimum mechanism, in which “*a mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon.*” (GLENNAN 2017, 17); and (c) Bechtel and Abrahamsen argue that “*A mechanism is a structure that performs a function in virtue of its components parts, component operations, and their organization*”. *Phenomena are responsible for the smooth functioning of the mechanism.* (BECHTEL and ABRAHAMSEN 2005, 423).

However, although there is not a consensus about the definition of mechanism, there are different aspects that are included in all these approaches. In this paper, I will analyze these aspects and propose a new mechanistic approach that will be able to address all them avoiding the problematic issues of classical mechanistic accounts.

## 2. THE MECHANISTIC PROPOSAL

It can be said that, in general terms, the different mechanistic proposals share the following four aspects, most of them already described in the foundational paper “*Thinking about Mechanisms*” (MACHAMER *et al.* 2000): First, mechanisms are formed by identifiable *entities* with certain stable and physically detectable characteristics that can be used to intervene in other processes, in addition to being physiologically plausible and relevant to explain the phenomenon under study (CRAVER 2006, 371). Second,



In addition to entities, second, mechanisms are formed by *activities* (MACHAMER *et al.* 2000), *interactions* (GLENNAN 1996), or *operations* (BECHTEL and ABRAHAMSEN 2005). Despite being intuitively the same concept, MCD underline an identical ontological consideration between activities and entities, which although still defended by one of its main developers, Carl Craver (2007, 64, 171, 177, 189), (2013, 7) is also criticized by authors like Kuhlmann and Glennan (2014) and Glennan (2017, 21). Third, although there is no consensus on how causality should be understood, it is agreed that mechanisms play a causal role in the occurrence of the phenomenon. It can be argued that the causal approach to mechanicism was firstly dominated by Wesley Salmon's (SALMON 1994) and Phil Dowe's (DOWE 2000) approach to the conserved quantities, and is now in the line of the manipulability theories of causation, such as the account developed by James Woodward (2003). However, there are also other causal approaches such as that derived from Glennan's own mechanicism (GLENNAN 1996) or Craver's constitutive approach. Fourth, the *organization* to which the mechanistic approach refers to typically corresponds to the hierarchical structure of many of the systems studied by scientists. Each component of a mechanism can also be itself a mechanism susceptible of being broken down into further mechanisms, which introduces the idea of *levels of organization* into the mechanistic framework. (MACHAMER *et al.* 2000, 13)

The idea of level of organization is one of the core notions for scientific research, since the choice of level depends on the entities, the actions that these entities carry out and the causal relationships among them. For instance, Craver (2007, 171) chooses the levels of mechanisms as the most valid option for carrying out an explanation, rejecting the classical view proposed by Oppenheim and Putnam (1958, 9) and the "refined" version of Wimsatt (1976, 253), that is, elementary particles, atoms, molecules, cells, multicellular living beings and social groups. A mechanism level is chosen, not only because of its entities, but also because of its activities. In other words, the election of the level or organization is function-oriented (cf. GARSON 2013).

However, Craver's decomposition of the levels gives rise to some difficulties. As noticed above, for Craver, a single-level element can be recursively broken down into sub-elements. However, it cannot be said that the resulting hierarchies are comparable, so that the sub-levels are out of sync. Therefore, as Craver and Bechtel point out, the size of the entities involved is not important, as a virus and a human may be at the same level of mechanism and interact with each other. From this contextual definition of mechanism, they attempt to explain downward causation without top-down

causes (CRAVER and BECHTEL 2006, 556) and exclude from their analysis strongly emerging properties on the basis that they are emergent only with respect to a different level of characterization. However, most mechanisticists advocate a mechanistic or organizational emergency in which superior properties are not directly derivable from those of their components because they are organization-dependent (CRAVER and TABERY 2017).

Despite the usefulness and ubiquity of the mechanistic explanation, I will argue that mechanism-level characterization is responsible for the difficulties cited above. In this paper, I will show how the notion of mechanism can be complemented with a new proposal that interprets the levels of organization of the matter as intervals of quasi-decompositionality (ICD), which will allow us to derive activities from the structure of the entities involved. This implies to make an *arbitrary but not capricious* choice of organizational level, explaining interlevel causation as merely apparent and avoiding the idea of emergent properties.

### 3. INTERVALS OF QUASI-DECOMPOSITIONALITY

I propose to depart from the idea of *nearly-decomposable systems* introduced by Simon in 1962. According to this proposal, certain systems can be split according to an arbitrary value  $\varepsilon$  into subsystems that are influenced only in an aggregative way. Unlike Simon's, my proposal considers the impossibility of distinguishing types of partitions according to their mutual influence, because no matter how small the influence between two partitions of a system is, it will never be null and void, meaning that the main premise in his account is never met. Even in the aggregated systems considered by Wimsatt (1997, S375), it can be found that the resulting properties assume more than mere accumulation. This can be seen by the classical example of a making a pile of sand, as not only does the mass and volume of the pile change, but the degree of compaction changes as well, therefore the density changes and properties such as fluidity *appear*. Therefore, I propose to eliminate the possibility of partitioning nature into decomposable elements, understood as those that are not influenced, and to consider that every material is a continuum in the purest style of (WEISS 1970, 7).

From this starting-point, properly justified in Cáceres and Saborido (2017; 2018), it can be stated that, if instead of taking a single criterion of decompositionality  $\varepsilon$  we take several criteria, it is possible to quasi-decompose the continuous material in the levels of organization of the matter (longitudinal intervals). In this way, one can decide on a value for  $\varepsilon$  that will provide us with a

fundamental or microscopic *aproblematic*<sup>2</sup> level F that delimit the partitions at the most basic level. Following Simon, this value could be based on the energy that holds matter together. For example, we can consider as a fundamental level the one that delimits the stability of the atomic nucleus ciphered around 8 MeV, a sufficiently high value to consider this level as “aproblematic”. With lower values, we could take the energy figures that delimit the chemical bond, especially the covalent<sup>3</sup>, so that we would have defined a second quasi-decompositionality interval (ICD) that would determine the molecules. Below this energy level would be the molecules and above this the associations of molecules. However, this choice is not without problems, as there are covalent bonds with a binding energy almost equal to that of some molecular interactions<sup>4</sup>. This only underlines the fact that quasi-decompositionality is ubiquitous and that the choice of values in  $\epsilon$  is essentially heuristic. Thus, the same energetic criterion allows us to continue quasi-decomposing matter in successive mesoscopic nested levels ( $m_i$ ) in until we reach a macroscopic level M that would encompass all matter. In this arbitrary but not capricious way, we can establish levels as quasi-decompositionality intervals equivalent to the classic atomic, molecular, cellular, organism, population and ecosystem levels, which partition nature into atoms, molecules, cells, organisms, populations, etc. This approach could provide the robustness that Glennan calls for when he says that *“parts of the mechanisms must have a kind of robustness and reality other than their place within that mechanism. In principle, it should be possible to take the part of the mechanism out and consider its properties in another context”* (GLENNAN 1996, 53).

Once a criterion has been established to decompose nature, we must select the ICD of our interest according to the phenomenon under study.

<sup>2</sup> Level F is the level of the fundamental particles described by Physics, a level that is usually considered to be the ultimate level that constitutes the basis for ontological reductionism. The Standard Particle Model considers that there are two types of particles, leptons and quarks, that lack internal structure, so they are considered truly fundamental particles. However, the stability of the atomic nucleus is so great that it can be considered as aproblematic level when it comes to its difficult decomposition. In addition to MDC, who considered this level to be “aproblematic”, Glennan (1996, 50) also referred to this fundamental level and has worked on the compatibility of the quantum level with the mechanistic explanations (KUHLMANN and GLENNAN 2014). From Physics, Roger Penrose (1991, 374-5) clarifies the difference between the quantum and the classical level, explicitly pointing out that the difficulties of current Physics do not really pose a challenge to the reductionist approach.

<sup>3</sup> Among them would be the ionization energies, so we could take the ions as an intermediate level.

<sup>4</sup> There are tremendously disparate covalent bond energies between them, for example, the  $E_{O-H}$  is 500 KJ/mol while the  $E_{F-F}$  is 160 KJ/mol. However, the energy of the hydrogen bond  $E_{F-H...F}$ , a non-molecular bond, is 155 KJ/mol.

Take, for example, the nutrition of vegetables. We can study this phenomenon from different perspectives. We can, for example, evaluate the energy flows from the sun and how they affect a forest, how they are transformed into plant matter, how they affect animal life, the climate, etc. We can also focus the study on how the light is captured by the plants through the leaves, how the flow of water and salts ascends through the xylem and how the sap is then distributed through the phloem, how the stomas of the leaves serve as an air and water vapour exchange zone, etc. But our interest may lie in the capture of energy from photons by protein photosystems, such as tilacoid molecules, that follow a sequence of oxidations and reductions that channel this energy, and how it is used to convert carbon dioxide and water into organic matter. Ecology, plant physiology or molecular biology will be the sciences responsible for carrying out these investigations.

Once the ICDs and their elements have been defined, a specific science will describe them structurally and functionally according to its methodology. Thus, ecology will not even distinguish species or will only distinguish them in terms of flow values, plant physiology will focus on the individual organism, and molecular biology on the chloroplast. Each field of science will develop its own methodology and language according to its own purposes, taking only those inputs and outputs it considers necessary and establishing a model system that will simplify reality heuristically. In this particular example, the concept of photosynthesis itself has a meaning related to matter and energy flows for ecology, fluids in plant physiology, and electrons in molecular biology. That is, when ecologists, physiologists and molecular biologists talk about photosynthesis, they mean different things. This is amplified as we ascend in level so that the observed generalizations will be equally inaccurate although they may be explanatory and predictive.

Each of the above sciences goes one step further in simplification. For example, molecular biology treats chloroplast as if it were what Bertalanffy calls an “equifinal state” (BERTALANFFY 1976, 147), i.e. it does not consider concrete chloroplasts, but rather idealizations of chloroplast formed by a complex of a priori selected interacting elements. These sciences will take this partition as their object of study, considering only the component parts it deems appropriate to explain what is of its interest. Thus, ecology will consider ecosystems and individuals and their functioning, physiology will consider cells, and molecular biology will consider atoms and ions. From this, we can infer that each science, in addition to its own  $\epsilon$ , uses as sub-quasi-decompositionality criteria ( $\epsilon_i^{\text{sub}}$ ) those of the sciences located at lower levels. Consequently, the sciences that investigate mesoscopic levels consider the immediately superior levels, and in a similar way they take the criteria of quasi-decompositionality of the sciences of the superior

levels as criteria of supra-quasi-decompositionality ( $\epsilon_j^{\text{supra}}$ ). Thus, with the choice of the levels at which a science has secondary competence, what can be called the extended quasi-decompositionality range  $IA_{(i,j)} = [\epsilon_i^{\text{sub}}, \epsilon_j^{\text{supra}}]$  is established.

This modeling, i.e. the description of a quasi-decompositionality interval with respect to the properties of that interval according to the science methodology of its level, implies an important simplification. A given level can be seen as a box to be opened by the science in charge of its study. In doing so, scientists will see that its interior is occupied by other enclosed boxes made up of lower-level systems, which are also studied by another science. Each box contains new boxes that correspond to another level and another science. This nesting of boxes would end at level F below and level M above. Therefore, it is inevitable that there will always be a loss of information that will have an important influence on the observation of regularities and therefore on the establishment of laws. The loss comes from taking as “real” values from the biases introduced when selecting the level, using a methodology, simplifying the previous interval, considering decomposability, etc. In addition, not all the variables that form part of a subsystem are used to describe them, but only those detected or considered relevant.

#### 4. MECHANICAL SYSTEMS

From this approach, only F-level entities have non-decomposable properties, while entities at all other levels will have a set of properties resulting from the simplification made when making the level choice. Therefore, an F-level entity will carry out actions based on their non-decomposable properties are ultimately due to their nature.

I label the result of the simplification made after choosing an ICD as a *system-entity*. This simplification involves not only the structural characterization, but also the selection of certain properties relevant to the explanation, which I call *level properties*. The system-entities are only arbitrarily stable, contrary to the idea of the classical entity of mechanicism. For example, following the example of photosynthesis, once the molecular level has been chosen, we can characterize RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) as the protein that carries out the first reaction of atmospheric CO<sub>2</sub> fixation, that is, the enzyme that initiates the transformation of inorganic to organic matter. We do not characterize this protein by its structure, which is very variable in all autotrophs that have it, but by some of its functions. In particular, we do it based on its carboxylase action, which initiates the dark phase of photosynthesis, and because of its oxygenase function, opposite to the

previous one, and which reduces the efficiency of photosynthesis (NAVARRO 2005, 5-9). Therefore, the RuBisCO is a modeling of a molecule that has some level properties that help us to explain the phenomenon that interests us, that is, photosynthesis. Therefore, the RuBisCO is a system-entity with certain level properties. Both its structure and properties are fully deductible from the lower level, despite the fact that they appear to be emergent properties.

The system-entities interact with others of the same level according to their level properties, configuring a mechanism that, seen from a higher perspective, supposes a system-entity of the next level.

In this way, we can define mechanosystem as follows:

*A mechanosystem (MS) is a set of system-entities (S-E) belonging to a quasi-decompositional interval (ICD) whose organization is due exclusively to the properties (P) of its components.*

$$MS = df( ICD, S-E, P)$$

Where:

- *the ICD is chosen according to the interest of the observer*
- *S-E is an entity-system and, at the same time, a mechanism of a lower ICD*
- *P are selected level properties of the previous ICD*

The particles of the fundamental level describe salmonian world lines<sup>5</sup>, and when they interact with those of other particles of the same level they exchange preserved quantities. Each nexus involves a causal interaction. Thus, the Salmonian view of world lines with a non-zero value of a conserved quantity and the intersection of these forming world networks is valid only for the fundamental level. As we ascend the level and speak, for example, of a molecule of water, we see how its atoms describe lines of world so that from a higher ICD it is perceived as a single thicker line, which will intersect with others in non-punctual nodes. As we move up more levels, the simplifications made in determining the mechanosystems introduce errors that accumulate making it difficult to detect the nodes. A set of water molecules together with a set of certain mineral salts can be seen as a concrete entity-system —the sap— that, at a higher level of analysis, serves to explain the ascension of nutrients by the capillarity of the trachea of the xylem. Similarly, the property “capillarity” is only observable at this level and, although it can be explained from the properties of its component elements, it is more easily distinguished by the properties of the level from which it is analysed, such as surface tension or density. Again, it can be said that this level property seems to emerge from

<sup>5</sup> A world line is the collection of points in a spacetime (Minkowski) diagram that represents the history of an object." (DOWE 2000, 90).

its components as a qualitative novelty, but it is actually nothing more than a phenomenon that can be explained in terms of the underlying level properties.

An interlevel observer will see that a fundamental  $F_0$  situation will be caused by another previous  $F_{-1}$ , its corresponding mesoscopic  $m_0$  will be caused by  $m_{-1}$  and the upper macroscopic  $M_0$  will be caused by a previous  $M_{-1}$ . When *zooming* in, the viewer cannot encompass the entire macroscopic phenomenon so he focuses on some aspect of the lower level, losing perspective. The ecologist is interested in balances, the physiologist in flows, the molecular biologist in chemical reactions. However, interactions take place exclusively at the fundamental level and therefore causality exists only at this level. For instance, the RuBisCO does not come into contact with  $CO_2$  as a whole, but only certain atoms of certain amino acids in that protein do so, meaning that such interaction is only real at the atomic level. This gives us an idea of the arbitrariness of the choice of ICDs, because even if we choose elements of the same level, we observe that there is no real causal interaction between them.

Hence, from this perspective the interaction between elements of different sizes proposed by Craver (CRAVER and BECHTEL 2006, 456-7) is problematic. It is true that scientists speak this way when referring to molecular mechanisms, but making a leap of at most one level. An explanation of the catalytic reaction of RuBisCO and  $CO_2$  includes expressions such as “*binding of RuBisCO and  $CO_2$  to the previously activated catalytic site*”, “*formation of 2,3-enediol by C3 deprotonation*”, “*nucleophilic attack of  $CO_2$  in C2*”, “*hydration in C3*” and “*deprotonation at C3-O level, which initiates the breakdown of the C2-C3 bond leading to two 3-phospho-D-glycerate molecules*” (NAVARRO 2005, 6).

When all the levels are observed at the same time, it is intuitive to analyze diachronic aspects as if they were synchronic, avoiding the differential of elapsed time  $dt$  and mistaking the causal explanation for the internal identity relationship. In this way, we run the risk of concluding that  $F_{-1}$  causes  $M_0$  and  $M_0$  causes  $F_{+1}$ , generating a false sense of upward and downward causality which, in reality, this is only apparent.

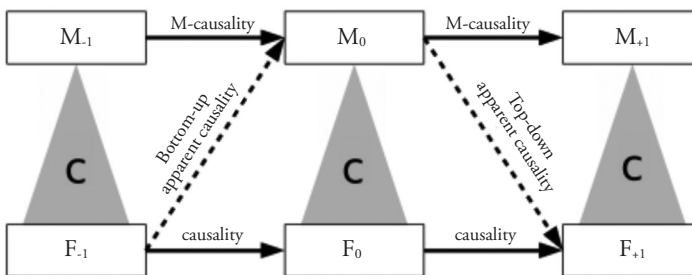


Figura 1: Apparent interlevel causality.

Thus, the nucleophilic attack of  $\text{CO}_2$  leads to an electronic rearrangement, which seen from a molecular perspective, is the breaking of a molecule into two. If we ignore  $dt$  we could believe that the nucleophilic attack is the cause of the formation of the two molecules, understanding therefore that there is an ascending causality. It is also common to express in the form of downward causation that RuBisCO maintains its active sites, when it really depends on the sequence of the amino acid's atoms and their interaction with the dissolving water.

If we consider, as this approach does, that the world is, ontologically, a material continuum, a criterion for the choice of mechanosystems should be defined. The first thing to do would be to determine the quasi-decompositionality interval. In principle, each science will choose as ICD the one that is part of its domain. Ecology, physiology and molecular biology will choose their respective ICDs. Once these levels have been established, it would be necessary to define which of the S-E's at that level are the ones that form part of the mechanism and, therefore, describe certain relevant level properties. Here we find two possible strategies: a) the elements are chosen, their properties are analyzed and the phenomena that occur are observed or b) the phenomena are chosen and the components and properties involved in their implementation are studied. Again, the decision between the two options is purely heuristic, being in general choice b) the most widely used in science. The recursive search for lower levels builds a nested set of mechanosystems. However, this top-down analysis from the phenomenon to the S-E has the disadvantage of excluding elements from the mechanism not only because of ignorance or errors, but also because of the quasi-decompositionality criterion that is assumed. A complementary bottom-up study from the S-E to the phenomenon can help to better understand the functioning of the mechanism and to better characterize the properties. In the example of photosynthesis, investigating the underlying components of this plant phenomenon resulted in the discovery of the RuBisCO and its function. But it was also discovered that the same enzyme is the one that catalyzes photorespiration, the "opposite" phenomenon to photosynthesis. A similar case is that of the so-called multifunctional or moonlighting proteins, which are those that perform different functions depending on their environment. A striking example is that of intrinsically denatured proteins (PINE) that are capable of changing their conformation, binding different ligands and performing different functions depending on environmental conditions (CUEVAS-VELÁZQUEZ and COVARRUBIAS-ROBLES 2011). In order to analyze this type of protein it is very effective to examine its behavior once it has been identified and biochemically described.



## 5. CONCLUSIONS

The consideration that mechanical systems<sup>6</sup> are quasi-decomposable allows us to develop an approach according to which levels of organization are heuristically deduced according to the entities of the system and their specific properties. The system-entities (S-E's) are interrelated, resulting in a multitude of outcomes some of which are subject to investigation. The organization of S-E's depends exclusively on their level properties, so the consideration of emergency, as J. S. Mill already pointed out, becomes an epistemic question (MILL 1843, 267). Depending on the level of the phenomenon which we are interested in, we will carry out our research by isolating the components that are important to us, thus describing a mechanism. Thus, for example, for ecology, photosynthesis is important because it leads to the production of organic matter and the beginning of food webs and the balances of matter and energy. Its entities would therefore be the plant populations. For plant physiology, photosynthesis is important because it is the form of nutrition of individual plants and governs their life. Its entities are therefore the organs involved, such as the leaves, roots, xylem and phloem beams and, in a lower step, the stomas of the leaves, the trabeculae of the conductive vessels, the absorbent hairs, etc. For cellular and molecular biology, photosynthesis is the fixation of inorganic carbon to organic matter. Its entities are the pigments of photosystems, light photons, enzymes such as RuBisCO and transformed molecules. And research can also be done at more basic levels, as is the case in recent studies of quantum effects in photosynthesis (THYRHAUG *et al.* 2018).

In conclusion, this proposal aims to enrich the neo-mechanicist view with a systemic, reductionist and non-eliminationist approach that understands the need to address the different levels of study from their own perspective with a practical aim.

<sup>6</sup> Although some authors claim that a mechanism can be a set of processes, which have a less stable configuration (ILLARI and WILLIAMSON 2012; GLENNAN 2017) the approach developed here interprets mechanisms as systems in which is possible to identify quasi-decomposable intervals.

## REFERENCES

- 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): 421-41.
- BERTALANFFY, L. VON. 1976, *Teoría general de los sistemas*, Fondo de Cultura Económica (ed.), Madrid.
- CÁCERES, E. & SABORIDO, C. 2017, "Intervalos de cuasi-descomponibilidad y propiedades emergentes", *Theoría* 32 (1): 89-108.
- CÁCERES, E. & SABORIDO, C. 2018, "¿Realmente mató la bacteria al coronel?: Perspectiva sistémica, causación internivélica e intervalos de cuasi-descomponibilidad en las explicaciones mecanísticas", *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia* 33 (1): 129-48.
- CRAVER, C. F. 2006, "When mechanistic models explain", *Synthese* 153: 355-76.
- CRAVER, C. F. 2007, *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*, Oxford University Press.
- CRAVER, C. F. 2013, "Functions and mechanisms: A perspectivalist view", *Functions: Selection and mechanisms*, 133-58. Springer.
- CRAVER, C. F. & BECHTEL, W. 2006, "Top-Down causation without top-down causes", *Biology and Philosophy* 22: 547-63.
- CRAVER, C. & TABERY, J. 2017, "Mechanisms in Science", *The Stanford Encyclopedia of Philosophy*, E. N. ZALTA (ed.), Spring 2017. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2017/entries/science-mechanisms/>.
- CUEVAS-VELÁZQUEZ, C. L. & COVARRUBIAS-ROBLES, A. A. 2011, "Las proteínas desordenadas y su función: una nueva forma de ver la estructura de las proteínas y la respuesta de las plantas al estrés", *TIP. Revista especializada en ciencias químico-biológicas* 14 (2): 97-105.
- DOWE, PH. 2000, *Physical Causation*, Cambridge: Cambridge University Press.
- GARSON, J. 2013, "The Functional Sense of Mechanism", *Philosophy of science* 80 (3): 317-33.
- GLENNAN, S. 1996, "Mechanisms and the Nature of Causation", *Erkenntnis* 44: 49-71.
- GLENNAN, S. 2002. "Rethinking mechanisms explanation". *Philosophy of Science* 69: S342-S353.
- GLENNAN, S. 2017. *The new mechanical philosophy*. Oxford University Press.
- ILLARI, PH. MCKAY & WILLIAMSON, J. 2012, "What is a mechanism? Thinking about mechanisms across the sciences", *European Journal for Philosophy of Science* 2 (1): 119-35.

- KUHLMANN, M. & GLENNAN, S. 2014, "On the relation between quantum mechanical and neo-mechanistic ontologies and explanatory strategies", *European Journal for Philosophy of Science* 4 (3): 337-59.
- MACHAMER, P.; LINDLEY D. & CRAVER, C. F. 2000, "Thinking About Mechanisms", *Philosophy of Science* 67 (1): 1-25.
- MILL, J. S. 1843, *A system of logic: ratiocinative and inductive*. [8th, 1872]. London: Longmans, Green, Reader, and Dyer.
- NAVARRO, J. V. M. 2005, "Contribución de residuos conservados de cisteína a la regulación redox del catabolismo de la Rubisco". PhD Thesis, Universitat de València.
- OPPENHEIM, P. & PUTNAM, H. 1958, "Unity of science as a working hypothesis", *Concepts, theories, and the mind-body problem, Minnesota Studies in the Philosophy of Science II*, H. FEIGL, M. SCRIVEN & G. MAXWELL (ed.), 3-36. Minneapolis: University of Minnesota Press.
- SALMON, W. C. 1994, "Causality without counterfactuals", *Philosophy of Science*, 297-312.
- THYRHAUG, E.; TEMPELAAR, R.; ALCOCER, M.J.C.; ŽÍDEK, K.; BÍNA, D.; KNOESTER, J.; JANSEN, T.L.C. & ZIGMANTAS, D. 2018, "Identification and characterization of diverse coherences in the Fenna-Matthews-Olson complex", *Nature chemistry*.
- WEISS, P. A. 1969, "The living system: determinism stratified", *Beyond Reductionism: New Perspectives in the Life Sciences*, A. Koestler & J. R. Smythies (ed.), 3-55. London: Hutchinson.
- WIMSATT, W. C. 1976, "Reductionism, Levels of Organization, and the Mind-Body Problem", *Consciousness and the Brain: A Scientific and Philosophical Inquiry*, I. SAVODNIK (ed.) & W. B. WEIMER (auth.), G. G. GLOBUS, G. MAXWELL. Springer US.
- WIMSATT, W. C. 1997, "Aggregativity: Reductive Heuristics for Finding Emergence", *Philosophy of Science* 64 (2): S372-84.
- WOODWARD, J. 2003, *Making things happen: A theory of causal explanation*, Oxford University Press.



JAVIER ANTA  
*LOGOS, Universitat de Barcelona*

# Indispensability and Effectiveness of Diagrams in Molecular Biology

Received: 6-7-2018 / Accepted: 12-3-2019

**Abstract:** In this paper I aim to defend a twofold thesis. On one hand, I will support, against Perini [7], the indispensability of diagrams when structurally complex biomolecules are concerned, since it is not possible to satisfactorily use linguistic-sentential representations at that domain. On the other hand, even when diagrams are dispensable I will defend that they will generally be more effective than other representations in encoding biomolecular knowledge, relying on Kulvicki-Shimojima's diagrammatic effectiveness thesis [4][11]. Finally, I will ground many epistemic virtues of biomolecular diagrams (understandability, explanatory power, prediction and hypothesis evaluation) on their cognitive-computational indispensability and their semantic-epistemic effectiveness.

**Keywords:** Molecular Biology, Diagrammatic Representation, Representational Indispensability.

## I. INTRODUCTION

The first thing you might notice when opening a biochemistry textbook is the astonishing amount of different visual resources that are employed, for instance schemas, flow charts, structural models, Haworth projections and so on. One could naively assume that the constant use of image-based representations, not just in textbooks but also in important biomolecular practices, only has an insignificant illustrative role as mere visual support of the main linguistically conveyed information. Otherwise and against this common prejudice, I am going to argue in this paper that not all but some visual formats, namely those which are

well-defined, syntactically-behaved and semantically-driven (from now on I will refer to them by the broad term of “diagrams”) plays a more than foundational role in the scientific discipline of molecular biology.

Although today is gaining much attention in the literature, the philosophical analysis of representational systems in special sciences and their semantic-epistemic implications is a relatively underdeveloped topic, with the outstanding exception of general and molecular biology [12], [7] or [1]. In this line of inquiry, Sheredos [10] expressed his curiosity on “why do biologists use too many diagrams?”. On my lights, the most plausible answer to this wide-scope question would be exactly the same than the one we could give to the more fine-grained “why might scientists prefer diagrammatic representations of information rather than, or in addition to, sentential ones?”. A tentative response to both the former and the latter questions (originally formulated by Bechtel and Abrahamsen [1]) will be sketched within the following pages. In the first section, I will argue against Perini [7] that sentential or linguistic formulas are not even possible for representing biomolecules having a high structural complexity, e.g. proteins at their crystallographic or quaternary structure level, and therefore diagrammatic representations would be indispensable within that broad domain. The thesis that diagrams are semantic and epistemically more effective than linguistic representation, in a general context and even when these latter vehicles are available, will be addressed in section 3. I will use Shimojima’s thesis of diagrammatic effectiveness and Kulvicki’s immediacy thesis (namely, diagrams are representationally effective because their relevant informational content can be highly available) to account for the observational advantages of biomolecular diagrams over formulas and sentences. In the last section, many epistemic virtues of diagrammatic reasoning in molecular biology (e.g. comprehensive, explanatory and evaluative advantages) will be assessed as intrinsically depending on the previously defended indispensability and effectiveness of these representational systems. Now, let’s start from the beginning.

## 2. INDISPENSABILITY OF DIAGRAMMATIC REPRESENTATIONS IN MOLECULAR BIOLOGY

First of all, it would be fair to claim that molecular biology is one the scientific area with more variety of representational systems for codifying information about their empirical domain, in a syntax-based and semantically-driven manner. Let us take the illustrative case of the biomolecule D-Glucose, and eight most frequent forms of representing it, as it is depicted in Figure 1. They range from the name “D-glucose”, its IUPAC nomenclature (first on the

left) wherein every piece of information about the molecule remains implicitly referred<sup>1</sup>, to one space-filling model of this biomolecule (first on the right), explicitly representing by graphical means a vast amount of physical and chemical properties, like van der Waals forces, which are encoded within the diameter of each ball. As one could learn from this eightfold representation, there do not exist a sharp distinction between fully diagrammatic non-diagrammatic, or fully sequential representational systems; the key differences are properly found in the particular mechanisms used for codifying information (for instance in Fisher projection, carbon atoms are represented by chemical symbol “C”, while in Haworth projection they are graphically encoded in the vertexes) about the 24 atoms of the D-glucose. It is worth mentioning that there also exist fixed semantic codes shared by many representational systems, like the CPK coloring (white for hydrogen, black for carbon...), that allow to systematically interpret certain properties and relations.

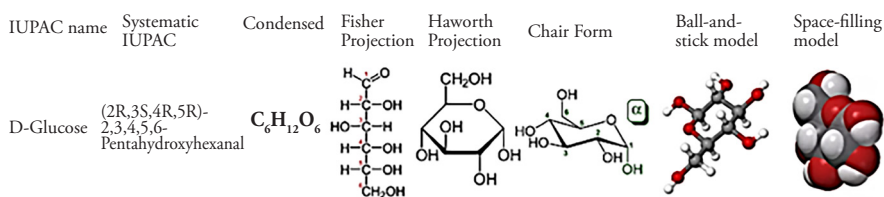


FIGURE 1. Representational manifold for the molecule D-glucose.

Laura Perini, one of the main philosophers devoted to assessing the representational and epistemic role of diagrams in biology, argues that the defining feature of diagrams (which properly demarcate them from linguistic representations) is the meaningfulness or significance of spatial properties and relations among the syntactically articulated graphical elements of the representing structure [5]. The syntactically-based and semantically-driven graphical behavior of diagrams is what differentiates diagrams from other kind of visual representations, like pictorial ones<sup>2</sup> [11]. For instance, one cannot graphically alter the relative position of the

<sup>1</sup> It would be highly controversial to assume proper names like “D-glucose” as representations if we consider representations as a sort of “homeomorphic relations” between the symbol and the represented phenomena. Here we are going to assume that representation relations are referred to different procedures of codifying information.

<sup>2</sup> Pictorial means of representation are those usually characterized as exploiting graphical resources but lacking of compositionality-systematicity. For instance, it cannot be possible to systematically articulate a new electron microscope photography “C” just from other EM images “A” and “B” (even when they depict the same protein, having the same content) precisely because they have neither well-defined syntactic rules nor compositional behavior.

bottom hydroxyl group “OH” attached to the anomeric or first carbon in the chair form (third on the right in Figure 1, green colored) without altering its semantic content: the lower “OH” opposed in the ring to the CH<sub>2</sub>OH group (indexed on carbon 5 and 6), which is known as a “trans” arrangement, graphically represents the specific alpha-anomeric structure of this molecules. In this sense, if this OH were just 1mm lower it would constitute a *meaningless* change, since this kind of structural diagrams are semantically sensitive not to the absolute location but to relative position (whether the OH is positioned “below” or “on the left”, as depicted in Figure 3) of their graphical elements<sup>3</sup>. This particular graphically-codified anomeric structure of the D-glucose, indexed in linguistic representations by an “a-” or “alpha-”, is only explicitly represented in the three diagrams on the right (Figure 1). Thus, diagrams are those representational systems wherein you can systematically change their meaning by syntactically manipulating graphical elements.

The fact is that diagrammatical alternatives in molecular biology are incredibly rich. Projective mechanisms of representation are particularly well-suited for codifying three-dimensional information in a schematic two-dimensional format: Fisher projective system make graphically explicit the organic or carbon-centered branching of biomolecules, grasping its chiral properties; the cyclic structure of carbohydrate become represented by means of Haworth representational system, which do not depict the actual but an idealized three-dimensional configuration of biomolecules (for that representational aim is effectively used its “chair form” projection). These diagrammatic mechanisms translate symbolic conventions of chemical notation, as used in condensed formulas, into sophisticated means of graphical descriptions of extensional structures.

In biochemistry, even the simplest object (for instance, the hydroxyl group “OH”) possess many structural subtleties. This plurality of biomolecular structures posit an important question for the purpose of this paper: could every piece of structural information about a molecule be linguistically codified or not? Perini assumes the idea that “analysis of diagrams shows that their content can be expressed with linguistic representations” [7, p. 257] or in other words:

- (1) **Diagrammatic Dispensability:** The informational content of a certain diagram or set of diagrams can be equivalently represented on a linguistic-sentential format.

Based on the notion of “computational equivalence” of Larkin and Simon, Perini took for granted that, although cognitively essential for

<sup>3</sup> I should thank an anonymous reviewer for suggesting me this point.



understanding certain complex phenomena, the content of biomolecular diagrams could be fully translated into serial or linguistic representation. She defends that sentential representations are always available, either as long conjunctive formulas or as a list of linguistic descriptions of each atom as the one we could find on a computer render software, and in this sense, any diagrammatic codification of the same data would be semantically dispensable satisfying (1). As it has just mentioned, Perini remarks that biomolecular diagrams are those kinds of representations which must be understood as “cognitively indispensable” (or “essential”, in her terms) for epistemic agents, not just to grasp complex information about those phenomena, but also to explain them:

The list of individual atomic coordinates would do little for a human in terms of understanding how these locations add up to the functional capacities of the complex. A serial representation of the positions of amino acids is readily available; it can be printed from the same electronically stored file of atomic coordinates which was used to make the diagram of the structure [5, p. 267]

I will support, in the forthcoming sections of this paper, Perini’s idea of biomolecular diagrams being cognitively indispensable (namely, epistemic agents needs diagrams for obtaining biomolecular knowledge) and explanatorily powerful; but, up to this point, I argue that (1) do not holds for the cases of codifying information about macromolecules with a high structural complexity; which is a foundational claim, since molecular biology and biochemistry are empirical domains wherein complex structures can be found everywhere. Linguistic-sentential representations are not always available in this domain. Then, one might have robust reasons to support the following thesis:

**(2) Diagrammatic Indispensability (at High Structural Complexity):**

The informational content of a determinate diagram or set of diagrams cannot be either computed nor equivalently represented on a sentential-linguistic format when this informational content possess a high level of structural complexity.

The motivation underlying (2) is not just that the information contained in a sentential representation of a complex macromolecule cannot be cognitively processed, but moreover, that this information cannot be (computationally) processed at all. Then, this would become a problem about the general computational impossibility (being human cognition a particular kind of computation) of processing such amount of information contained in

a single formula describing complex macromolecules at the atomic level. As Perini suggested in the above quote, the only actual serial formula of this kind is just a string of software code (a code sample is depicted on the right of image 2) whose unique semantic role is rendering the macromolecular diagram, which cannot be directly used for any epistemic activity.

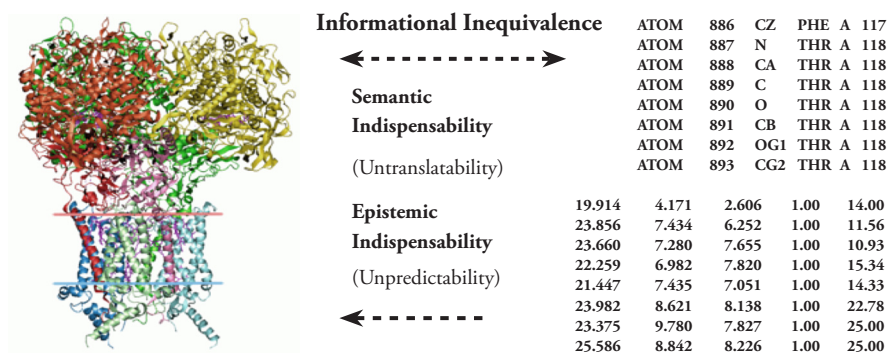


FIGURE 2. Representing the high structural complexity of human deoxyhemoglobin, 4hhb. (Left) Crystal structural three-dimensional model from the Protein Data Bank, RCSB. (Right) Software code specifying a list of atomic values required exclusively for rendering the structural model.

In which sense should be understood structural complexity in molecular biology? Let us assume as a paradigmatic case of complex macromolecule the human deoxyhemoglobin, which is a protein containing a polypeptide chain of 154 amino acid (at the monomeric level) or 1579 atoms (at the chemical level). (i) Every amino acid encompasses atomic properties and relations among atoms; (ii) the primary structure or polypeptide chain of the deoxyhaemoglobin encompasses molecular properties, intramolecular and intermolecular relations and higher-order relations among atoms; (iii) its secondary structure (α-helix, for instance) entails higher-order intramolecular and intermolecular relations and exponential-order relations among atoms. We reach the critical structural complexity threshold at the level of (iv) deoxyhemoglobin's tertiary (as well as quaternary) structure, wherein the so-called computational complexity of a possible symbolic description became logarithmic: which means that exist an asymptotic approach to the use of infinite computational resources to solve the task.

For achieving this purpose in a sentential manner, it would not suffice with the chemical formula containing every atom within deoxyhemoglobin; notice that this kind of representation would have no structural information at all. Otherwise, it would be required one hypothetical formal-mathematical language able to express every single relational value, (e.g. complete set of atomic coordinates, relative positions, angular separation, etc.) required

for an exhaustive description of the protein or any other similarly complex macromolecule. Rejecting Perini's argument, I have just argued above that that no software code (see image 2) would ever satisfy those requirements. My claim is that such hypothetical symbolic apparatus would also suffer from the previously mentioned computational impossibility. At this point of structural complexity, only diagrammatic devices can computationally perform the representational work, and Perini is aware of it, in spite of her supporting (1):

What makes structural models so important is the fact that amino acid chains do not simply stretch out in a line. They wind around in complicated formations. This means that side groups on amino acids that are very far from one another in terms of position on the amino acid chain might be located right next to one another in the protein [6, p. 267]

The main reason of why this computational impossibility is not the case for crystallographic models or structural diagrams is because of their idiosyncratic information-encoding mechanisms: they unload complex structural information extensionally codified across the representation, while in the case of sentential representations the intensional codification remain fully implicit in the symbols, involving an overload of computational resources to perform the same descriptive task. Up to this point, we should remark that even a sophisticated structural-diagrammatic model won't be able to represent all possible spatial relations within a determinate macromolecule, since it would rapidly exceed any conceivable set of computational resources. Traditional biomolecular diagrammatic systems (see Figure 1) are representationally constrained limited to a single level of organization, i.e. ball-and-stick models to atomic structures, ribbon diagrams to polymeric units, and so on. Interestingly, new software-enhanced structural diagrams, satisfactorily encoding multilevel (atomic, molecular and polymeric) information, have been developed recently.

The most important databases of macromolecules, for instance the Protein Data Bank (PDB) or Proteopedia, employ complex 3D or stereographic (so they are not "static" or "printed") diagrammatic-structural representation<sup>4</sup> of proteins and other complexly structured biomolecules to organize relevant and novel information about the field. PDB diagrammatic-models, like the one depicted in Figure 2, encode information about several levels of biomolecular information in an interactive way: you can select whether to visualize (i)

<sup>4</sup> This 3D representation (2D projected onto the screen or the paper) could be stereographically rotated, so you could visualize the back of the protein in Figure 2.

its tertiary-secondary structure (protein strand, helix structures), (ii) its amino acid chain (primary structure) or even (iii) their underlying atomic architecture. These recent software-based diagrammatic representations also allow to overlap its protein-level interface (e.g. Figure 2) with its underlying atomic constitution, so you can visualize the exact location of oxygen atom within some helix structure.

Of course, there also exist databases of protein and DNA sequences, being NCBI one of most used, but one important fact within the field of bioinformatics is that structural-diagrammatic model already contains (and make explicit) its sequential or primary structure and not the other way around. This asymmetry is foundational for a second biomolecular practice that has recently gained much interdisciplinary attention, which is known as PSP<sup>5</sup> (Protein Structure Prediction) and consists on inferring highly complex (tertiary-quaternary) structure from its sequence or sentential representation. For carrying PSP, it is necessary continuous work and sharing computational resources in a worldwide scientific community (named CAMEO3D) exclusively devoted to the performance of these structural predictions. Even with both huge international cooperative effort and incredibly sophisticated new techniques, the current maximal accuracy for predicting *just the secondary structure* of a protein (notice that its biological functions rely on its tertiary-quaternary structures) from its primary structure is about 80%. Therefore, it would be required an asymptotically infinite quantity of computational resources for predicting higher protein structures from sequences of polypeptides.

...

Up to this point, we can firstly answer the Bechtel-Abrahamsen question “why might scientists prefer diagrammatic representations of information rather than, or in addition to, sentential ones?” in the following term: when information about highly complex biomolecular structures is concerned, molecular biologist can only have structural models or diagrammatic representations (even when they are restricted to a single level of organization) since *there are no* equally informative sentential-formulaic encodings of these macromolecules available. Furthermore, as I have defended: *there cannot be* linguistic representations at those levels of structural complexity. In the following section, I will try to give an answer

<sup>5</sup> I should thank Nuño de la Rosa for reporting me the fact that it cannot be predicted the tertiary or quaternary structure of any protein from its primary or sequential structure by any current software.

to the above question for general biomolecular contexts of lower structural complexity.

### 3. THE REASONABLY EFFECTIVENESS OF BIOMOLECULAR DIAGRAMMS

It has been argued during the previous section that there are no (and moreover, there cannot be) properly linguistic or sentential representations of complexly structured biomolecules; therefore diagrammatic ones or structural models stand as indispensable means of codifying biomolecular information in the domain of high structural complexity. In this section I will argue that, additionally to their indispensability, diagrammatic systems are much more (i) representational and (ii) inferentially effective than formula-based means, not just in the not narrow domain of complex biomolecules, but in the general disciplinary context of molecular biology. Secondly, I will also try to offer one suitable naturalistic explanation of this reasonable effectiveness of biomolecular diagrams.

Shimojima formulated the question of why some diagrams are much more representationally effective than sentential in certain scientific practice context. His answer to that question is that one specific diagram would be representationally effective when its codification of information entails certain “observational advantages”, like the possibility of reading-off its content without performing inferences. This is equivalent to what Kulvicki [4] calls the “immediacy of information” due to the extractability, syntactic salience and semantic salience. For instance, a chair form projection of a glucose molecule would be highly effective in encoding enantiomeric and anomeric structure of this molecule if one could tell *a*-D-glucose from *b*-D-glucose by simply reading-off the biomolecular (semantically salient, in Kulvicki’s terms [4]) content of diagrams, as shows Figure 3. At least, this projective format would much more effective when concerning anomeric structures than other diagrammatic or sequential representational systems, like a fisher projection or a condensed formula respectively, simply because these two cannot representationally distinguish between *a*-D-glucose and *b*-D-glucose while a chair form projection can.

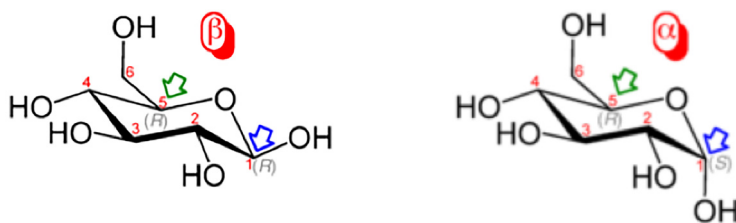


FIGURE 3. Anomeric structure of the D-glucose diagrammatically represented.

In this sense, Shimojima would claim that a chair form projection of a *a*-D-glucose can be used to read off its anomeric properties, which constitute a “observational advantage” against formulaic representations like the IUPAC name “*a*-D-glucose”. In words of Kulvicki [4], the anomeric information would be immediately available from the diagrammatic representation precisely because of its syntactic-semantic salience. Even a 6-years old child could tell two anomericly different D glucoses apart from their chair form projections (by simply reading their informational content off from the below/right relative position of OH), but obviously she cannot do the same from their IUPAC names. Then, it can be said that biomolecular diagrams are representationally effective in many levels of biomolecular abstraction (i.e. atomic, monomeric or polymeric) because of the semantic-syntactic immediacy of their graphically-encoded informational content: this is “Kulvicki-Shimojima’s effectiveness thesis”, henceforth KSET.

Up to this point, it would be important to make the distinction, firstly stated by Larkin and Simons [5], between two representations being “informationally equivalent” (namely, the two implicitly or explicitly contain exactly the same information) and two representations being “computational” or “inferentially equivalent”. This latter concept refers to two representations with the same inferential power, wherein the same kind of inferences can be performed on both. So, in the previous case of the “*a*-D-glucose” and its chair form projection, one could notice that the two representations are informationally equivalent but inferentially inequivalent, since a molecular biologist could effectively infer much more information from the projective diagram (e.g. where are located the binding atoms within the D-glucose molecule, how would it behave within an acid environment, etc.) than from its IUPAC name.

In the line of Suárez’s [6] inferentialism, Bueno [2] developed the view that diagrams could be treated not just as syntactically-structured representations but moreover as epistemically significant “inferential devices”. Then, the key difference between diagram-based inferences and formula-based inferences is that the former ones are (perceptually) carried immediately on the representational symbols, while the latter ones are (cognitively) performed mediately based on the meaning of the representation. As far as these two clearly differentiated inferential behaviors would determine different ways of obtaining biomolecular knowledge, we should also distinguish between the semantic and the epistemic dimension of KSET:

- (3) **Semantic Effectiveness of Diagrams:** The semantic or representational effectiveness of a diagrammatic system is given by the (computational) exploitation of geometric-topological resources for explicitly codifying and extracting relevant information.

- (4) **Epistemic Effectiveness of Diagram:** The epistemic effectiveness or inferential power of a diagrammatic system is given by the (algorithmic/cognitive) exploitation of geometric-topological resources for processing relevant information.

Departing from this two-fold thesis, we can offer an answer to the question of why diagrams are so (semantically) effective and have so many observationally advantages within molecular biology. The essential idea is that by graphically exploiting geometric-topological resources (e.g. projective spaces, invariant connections and shapes, geometrical configurations, etc.) in diagrammatic means it could be possible to encode relevant structural information about biomolecules without inflating syntax exponential or logarithmically, as in the case of sentential representations, or losing operational rigor, as in the case of pictorial means. Or in other terms: the semantic effectiveness of one particular representational system within a certain domain (e.g. structural models and diagrams in molecular biology) could be understood by means of representational mechanisms making relevant information more available. By relevant information in the biomolecular domain I specifically mean the minimal set of values that determine the value of any biochemical property of the macromolecule. It would be easy to demonstrate how these foundational biomolecular values are extensional values, wherein the “extensional” here refer to any space-constrained property: atomic position, distance between functional groups, angular separation and so on. In the end, the astonishing effectiveness of diagrammatic representations intrinsically depends on the intrinsic spatial character of biomolecular phenomena at any level of organization (i.e. pentagonal benzene, helicoidal secondary structure); fact that can be satisfactorily exploited by well-suited graphical mechanisms.

As a brief parenthesis, it is worth mentioning that this application of the KSET (3) and (4) to the scientific domain of molecular biology could explain why diagrammatic indispensability, under the conditions specified in the previous section, holds. The fact the certain diagrams and some formulas might be homeomorphic with higher biomolecular structures and biochemical sequences, respectively, does not tell us any relevant nor explain the diagrammatic effectiveness thesis. Otherwise, it could be satisfactorily explained by the fact that the structural complexity of macromolecules is well-suited for being diagrammatically and extensionally codified on symbol: for instance, complexity of a protein's quaternary structure become codified in the graphical entanglement of a structural model. While on the other hand, it could be sequentially-intentionally codified in the syntactic complexity (namely, how the symbols like “OH” or “C” are sententially articulated) of the correspondent language, which is precisely what underlies the computational overload mentioned before.



So far, we have been referring to the representational effectiveness of diagrams (3) in molecular biology; but this kind of visual representations also play a decisive and foundational epistemic role in this scientific field, as it will be fleshed out in the next section. The epistemic effectiveness of diagrams (or in other words, the semantic-pragmatic necessity of diagrams for obtaining and manipulating biomolecular knowledge), as formulated à la Shimojima in (4), can be based on the same fact than its representational effectiveness: the suitability of geometric-topological resources for encoding and processing relevant biomolecular information. That is precisely the reason why diagrams, being semantic and epistemically superior to sentential representation, are constantly used in actual practices of molecular biology: they are required for (i) facilitating compressibility and understandability of complex objects-mechanism in pedagogical contexts, (ii) many different kinds of biomolecular explanations (structural-functional, dynamical-mechanistic or even topological), (iii) testing novel hypothesis and discovering new phenomena, and (iv) predicting biomolecular events. Such rich manifold of epistemic virtues in biomolecular diagrammatic reasoning, grounded on its indispensability and effectiveness (as it has been defended in this paper), will be properly evaluated in the following section.

#### 4. EPISTEMIC VIRTUES OF DIAGRAMS IN BIOMOLECULAR REASONING

I have tried to show in the two previous sections that if molecular biologist rather use diagrammatic representations in their scientific practices is precisely because (i) in some occasions there are no sentential-linguistic options and (ii) in most occasions diagrams are more effective in codifying biomolecular information and knowledge than non-diagrammatic means. Nevertheless, I will argue in this section that diagrammatic representations are also superior to other representational apparatuses not just in describing macromolecules but also in explaining and understanding the biological function of certain chemical compounds or the mechanism underlying certain biochemical processes and mechanism, in assessing new hypotheses or even in predicting how biomolecules will behave under specific conditions.

#### **Visual-Diagrammatic Understanding**

One of the clearest epistemic advantages of diagrammatic representations over sentential formulations of biochemical information is the high degree of understandability or comprehensibility of the former over the later



format, which is not only important at the pedagogic level (it should be noticed that biochemical textbooks can be valued according to the quality of their illustrations [7]) but also in hypothesis-testing and explanation. With certainty, it might be claimed that molecular biology is one of the scientific disciplines whose epistemic agents dependent cognitively more on visual means (diagrams, schemas, pictures, etc.) of conveying relevant information. At this point, we can interpret our thesis of diagrammatic indispensability (2) not in general computational terms, as it was done in section 2, but moreover from a cognitive perspective: it would be impossible, even for a professional molecular biologist, to understand or fully grasp the informational content of any sequential-sentential representation of a complexly structured macromolecule, precisely because the resources for cognitively processing it exceed by far human capacities. Then, as Perini defends, diagrammatic formats in the biomolecular domain are cognitively essential or indispensable.

### **Functional-Structural Explanations**

The topic of diagrammatic explanations in empirical sciences is clearly underexplored. Perini [7] argued that this is mainly because since the deductive-nomological model of scientific explanations it has been assumed that only linguistic representations can be explanatorily relevant. Against this prejudice, one of the most relevant epistemic capacities of diagrams in general biology (but also in biochemistry and molecular biology) is, also according to Perini [7], to develop functional explanations by exploiting semantic-syntactically visual resources to graphically remark the functional relations existent among the parts of a certain molecule. I strongly agree with her in the thesis that diagrammatic systems are better suited than sentential representations in carrying those particular kinds of explanations, which might be regarded as a consequence of epistemic effectiveness (3) previously defended:

The unique capacity of visual representations to represent higher-order relations in virtue of visible relations holding among parts of the figure suits them to functional explanation in a way that representations with a linear format cannot match [7, p. 267]

One structural model or one schematic diagram, as explanans, could be functional-structurally explanatory by making salient certain features within the visual representation that encoded the relevant information about

the explanandum. Let us consider the illustrative case of the biomolecular mechanism underlying bronchodilation as *explanandum*: it could be possible to explain in a signal transduction context how binding interactions and the functional import of the receptor are mediated by electrostatic effect; and we do it by representing the macromolecule (cell-membrane spanning)  $\beta 2$  adrenoreceptor in a complex space-filling model or diagram. The proper explanatory power of the diagram as *explanans* depends externally on (i) a minimal background knowledge of molecular biology and enzymology (facilitating understanding of the diagram), and intrinsically on (ii) graphically reinforcing the relation between those functionally-relevant structures within the macromolecule, in this case by a chromatic codification of biomolecular electrostatic properties.

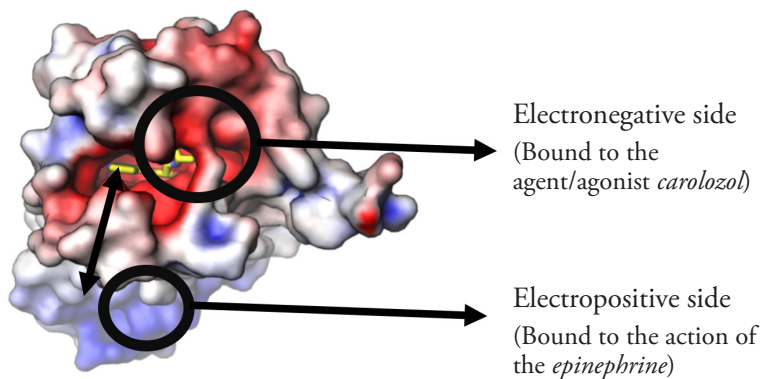


FIGURE 4. Explanation of the function of  $\beta 2$  adrenoreceptor in bronchodilation.

### Mechanistic-Dynamic Explanations

The explanatory power or epistemic effectiveness of diagrammatic representations is far from being constrained to the functional-structural kind of explanations. Another closely-related kind of diagrammatic explanation very frequent within the biomolecular context is what is known within the literature as “mechanistic explanation” [1], and when time is a key factor they are also called “dynamic explanation”. Sheredos [10] argues that, although it is perfectly possible to offer mechanistic-dynamic based on linguistic descriptions, diagrammatic representations of biomolecular mechanisms will always be both semantically and epistemically privileged against sentential representation. I would defend that the underlying generic reason for this privilege is precisely (3) and (4). A much more specific reason is that mechanistic explanations are usually displayed via causal chains or cause-effect sequences; in that sense, a sequential-formulaic (one-dimensional) representation could

only codify information about causal paths in a serial fashion<sup>6</sup>, while a diagrammatic (two/three-dimensional) representation would be able to encode data simultaneously about many possible causal chains within the particular mechanism. Another important observational advantage of diagram-based mechanistic explanation, particularly in metabolic pathways and enzymatic networks, is that graphical resources can be used for effectively representing those molecular structures changing during a biological process as well as those that remain invariant. For instance, we can use a metabolic pathway diagram (containing Haworth projections as sub-diagrammatic units) as the one depicted on image 5, for developing a properly mechanistic explanation of how pyruvate can be obtained from glucose by the enzyme-catalyzed biological reaction of glycolysis. A set of causal chains is diagrammatically depicted (the so-called EMP pathway) supported by a chromatic code, wherein red means changing structures, blue arrows indicate ATP consumption and yellow ones do the same with ATP production.

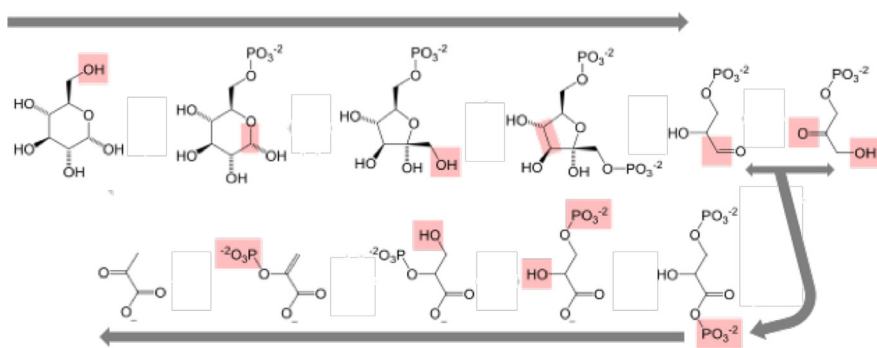


FIGURE 5. Dynamic explanation of glycolysis process based on Haworth projections.

## Hypothesis Testing and Discovering

As a conclusion of this section, it worth mentioning briefly the highly efficient use of diagrams in the evaluation of biomolecular. Bueno [6] reported a case wherein computer-generated structural models of a particular protein were statistically compared with structural data obtained from x-ray crystallographic measurement of the same protein. Due to a poor 63 % correlation between the empirical image and the structural model (created under the theoretical assumption that the spatial configuration of the protein's surface cannot change), a new hypothesis was introduced:

<sup>6</sup> Of course, you can use a formulaic representation to represent multiple causal path, but always in a serial way.

The researchers then returned to the theoretical image, and made a novel hypothesis. (...) given that the surface molecules have incomplete bonds, and are in contact with the environment, they tend to interact with it, thus changing the packing arrangement. (...) The researchers then changed the theoretical image by incorporating the assumption that surface reconstruction took place. The correlation now between the theoretical and the experimental images was 93%—a far more significant result [1, p. 663]

Due to the huge impact of the received view in the general philosophy of science, theoretical content and hypotheses were assumed to be only conveyed in symbolic-logic formulas or in ordinary language sentences; since the 90s, philosophical interest on real scientific practices lead to also regards non-linguistic formats, like the computer-rendered structural models of the example, as perfectly legitimate vehicles for abducting theoretical content. In this sense, I would strongly argue that (widely-conceived) diagrammatic activities should be regarded as the main form of creating, manipulating and rejecting biomolecular hypothesis. Thus, as the above case shows, diagrammatic procedures cannot be merely conceived as convenient semantic resources, but moreover as essential epistemic tools which also allow scientist to discover novel phenomena (i.e. rearrangement of surface configuration took place in organic crystals), and therefore as essential for the disciplinary development of molecular biology.

## 5. CONCLUSION

Summarizing, all along this paper it has been defended a twofold answer to the question of why might molecular biologist prefer diagrammatic representations of information rather than sentential ones. On one hand, I have argued against Perini [5] that it is not even possible to represent structurally complex biomolecules by means of sentential or formula-based vehicles precisely because it would be required (i) a logarithmic quantity of computational resources to uncodify the informational content of such semantic object or (ii) non-human cognitive abilities for an epistemic agent to be able to understand it. Then, diagrammatic forms of conveying information would be semantic and epistemically indispensable at this level of biomolecular structural complexity. On the other hand, I have also defended the thesis that, even when diagrammatic systems are dispensable, they are much more effective than linguistic representations in both codifying biomolecular information and in obtaining knowledge from that empirical domain. The reason for this semantic-epistemic effectiveness could be found in the “representational

immediacy” [4] of geometry-exploiting diagrammatic mechanisms (according to Shimojima [9]) while codifying relevant structural biomolecular data. Finally, I have pointed out that both the defended indispensability and effectiveness of this kind of syntactically sophisticated visual representation might satisfactorily explain its epistemic virtuosity in many real molecular biology practices: diagrammatic understanding enable to grasp complex biochemical concepts as well as make possible to carry functional-structural explanations, additionally they can also be used to develop other explanatory modalities, like mechanistic-dynamic explanations of enzymatic processes or topological explanations of how proteins unfold, as well as to evaluate novel hypotheses and discover new phenomena. In conclusion, diagrams should not be conceived as mere useful and illustrative tools for depicting molecules or biochemical reactions; moreover, these representational systems are foundational, essential and pivotal in a vast plurality of ways (e.g. cognitively, computationally, representationally, explanatorily, predictively...) to this scientific domain as indispensable and effective vehicle of biomolecular knowledge.

#### *Acknowledgements*

Firstly, I am really grateful to an anonymous reviewer for her quite useful comments. I should also thank Javier Suarez, Laura Nuño de la Rosa, Roger Deulofeu and Álvaro Ropero for fruitful discussions. This paper was originally presented during the “Eight Philosophy of Biology and Cognitive Sciences Research Workshop” at the Complutense University of Madrid (May 2018) and it was developed with the academic support of LOGOS (research group in analytic philosophy) and the reach group “Laws, explanation and realism in physical and biomedical sciences” (FFI2016-76799-P). The research crystalized in this paper was also economically funded by the Spanish Ministry of Education (FPU16/00774).

## REFERENCES

- [1] - BECHTEL, W & ABRAHAMSEN, A. 2005, "Explanation: a mechanist alternative", *Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 421-41.
- [2] - BUENO, O. 2012, "Styles of reasoning: A pluralist view", *Studies in History and Philosophy of Science Part A* 43 (4): 657-65.
- [3] - JONES, N. 2014, "Bowtie Structures, Pathway Diagrams, and Topological Explanation", *Erkenntnis* 79 (5):1135-55.
- [4] - KULVICKI, J. 2010, "Knowing with images: Medium and message", *Philosophy of Science* 77 (2): 295-313.
- [5] - LARKIN, J. & SIMON H. 1987, "Why a diagram is (sometimes) worth ten thousand words", *Cognitive Science* 11: 65-99.
- [6] - SUÁREZ, M. 2002, "An inferential conception of scientific representation", *Philosophy of Science* 71 (5): 767-79.
- [7] - PERINI, L. 2005, "Explanation in two dimensions: Diagrams and biological explanation", *Biology and Philosophy* 20 (2-3): 257-69.
- [8] - PERINI, L. 2012, "Form and Function: A Semiotic Analysis of Figures in Biology Textbooks", N. Anderson & M. Dietrich (ed.), *The Educated Eye Visual Culture and Pedagogy in the Life Sciences*, 235-54.
- [9] - PERINI, L. 2013, "Diagrams in Biology", *The Knowledge Engineering Review* 28 (3): 273-86.
- [10] - SHEREDOS, B.; BURNSTON, D.; ABRAHAMSEN, A. & BECHTEL, W. 2013, "Why Do Biologists Use So Many Diagrams?", *Philosophy of Science* 80 (5): 931-44.
- [11] - STAPLETON, G.; JAMNIK, M. & SHIMOJIMA, A. 2017, "What Makes an Effective Representation of Information: A Formal Account of Observational Advantages". *Journal of Logic, Language and Information* 26 (2): 143-77.
- [12] - TAYLOR, P. & BLUM, A. 1991, "Pictorial Representation in Biology", *Biology and Philosophy* 6: 125-34.

MARÍA J. FERREIRA RUIZ  
*CONICET, Universidad de Buenos Aires*

# How *not* to Resist the Natural Kind Talk in Biology

Received: 6-7-2018 / Accepted: 17-1-2019

**Abstract:** After the dawn of the traditional, essentialist view of natural kinds in contemporary philosophy (exacerbated in philosophy of biology by “population thinking”), non-essentialist cluster conceptions of natural kinds have been extensively supported and applied to numerous biological categories. However, salient philosophers have put forward two challenging arguments against cluster kind theories. I argue that, in both cases, discontent with a *cluster* conception of natural kinds is motivated by tacit and previous assumptions that can be challenged. I conclude that the concerns expressed in the objections do not make good reasons to resist natural kinds talk in biology *unless* one is willing to share such commitments and assumptions with respect to natural kinds. Ultimately, the discussion can be used to point out that our very expectations regarding natural kinds theories could use a rethink.

**Keywords:** natural kinds, essentialism, cluster kinds, Boyd, mechanisms.

## I. INTRODUCTION

ACCORDING TO A TRADITIONAL CONCEPTION in contemporary philosophy, natural kinds are defined by necessary and sufficient conditions, which constitute the kind’s essence (PUTNAM 1973; KRIPKE 1980). Essential properties are intrinsic features, typically, to be found in a thing’s micro-structure. Such properties are possessed by all and only the members of the kind, and they must explain why members share other superficial features. For the most interesting scientific cases, such a definition was hard to achieve, and this traditional conception received extensive criticism (see BIRD and TOBIN 2015).

In the philosophy of biology, the most widely discussed case of a natural kind is that of biological species (BOYD 1999; GRIFFITHS 1999; MILLIKAN 1999; WILSON *et al.* 2007). However, many authors concerned with species turned against the traditional conception of natural kinds. It was argued that an essentialist conception of species is precluded by or incompatible with the “population thinking” that characterizes modern Darwinian biology (SOBER 1980). Mayr argued that the Darwinian theory involves the replacement of an entire mode of thinking in biology, namely essentialism, with an opposite one, namely, population thinking. In his own words:

the populationist stresses the uniqueness of everything in the organic world [...]. All organisms and organic phenomena are composed of unique features and can be described collectively only in statistical terms. Individuals, or any kind of organic entities, form populations of which we can determine the arithmetic mean and the statistics of variation. Averages are merely statistical abstractions, only the individuals of which the population are composed have reality (MAYR 1976, 28-9)

The typologist (as Mayr calls it) would regard variation as an illusion, whereas the populationist would take the type to be an abstraction and take variation to be real. From the populationist perspective, then, any essentialist view of natural kinds would be unsuited for biological kinds.

A different conception of natural kinds—one that has been extensively applied in biology—is found in Richard Boyd’s seminal work. According to his Homeostatic Property Clusters theory (HPC, hereafter) kinds are not definable in terms of sets of necessary and sufficient conditions, but in terms of a cluster of shared properties and a set of mechanisms responsible for such clustering, both of which are conceived in an open-ended manner. Cluster views of natural kinds have been further developed over the years and they seem to be better prepared to account for classification practices in biology (e.g., to account for species and higher taxa, cell types, homologies, developmental modules, genes, life, or ecosystems).

Nonetheless, and despite the general acknowledgement of the merits of cluster views, compared to a traditional, essentialist conception, some philosophers believe that cluster views are too weak or flexible. More specifically, there are two important objections that, surprisingly, have not been properly analyzed and contested: (a) *the explanatory limitation objection* and (b) *the membership determination objection*. My aim in this paper is to dismantle these objections from a naturalistic, anti-essentialist point of view.

I will argue that in both cases, discontent with a cluster conception of natural kinds is motivated by tacit and previous assumptions with respect



to natural kinds, which involve the aim of a theory of natural kinds and expectations regarding scientific classification systems. I will challenge these assumptions, showing how they are unwarranted, and conclude that the concerns expressed in the objections do not make good reasons to resist natural kinds talk in biology *unless* one is willing to positively endorse such commitments. My considerations do not rely on peculiarities of any particular biological kind (species or higher taxa, cell types, genes, etc.), but on more general ideas about natural kinds and classifications. Importantly, I will not argue here that there are genuine natural kinds (biological or otherwise). This paper will not contribute to defend natural kind realism. Rather, I aim at undermining two ways of resisting natural kind talk in biology, namely, on the basis of these two objections.

## 2. NATURAL KINDS IN CONTEMPORARY PHILOSOPHY: MICRO-ESSENTIALISM AND PROPERTY CLUSTERS KINDS

The topic of natural kinds is one of the most classic topics in metaphysics and philosophy of science. It is a common belief that objects in nature can be classified into different kinds of things, on the one hand, and that natural scientific theories and disciplines talk about these kinds, on the other. Stances on the naturalness of kinds (are there genuine divisions in nature?) range from realism to conventionalism. A realist believes that genuine divisions exist independently of human practices, even if we reflect them incorrectly. In contrast, a conventionalist stands for the idea that all kinds are epistemic construals that reflect our needs and interests (be daily or scientific), with no existence of their own.

The traditional realist conception of natural kinds in contemporary philosophy is mainly due to Saul Kripke (1980) and Hilary Putnam (1973; 1975). According to their view, natural kinds are defined by a conjunction of individually necessary and jointly sufficient conditions, which constitute the kind's essence. Essential properties are intrinsic features, typically, to be found in the thing's micro-structure. Such properties are possessed by all and only the members of the kind, and they must explain why members share other superficial features. Essential properties also explain one of the most salient features of natural kind categories in the traditional view: their projectibility. Traditionally, natural kinds are categories used in explanation and prediction (sometimes expressed in the requirement that they must feature in scientific laws). A well-known example discussed by Kripke is gold. Having the atomic number 79 is an essential property of gold, which explains its

other microchemical or phenomenic properties. Something similar to gold in superficial respects (color, for example) but which lacks the property of having the atomic number 79, would not be gold. Putnam argues for similar micro-essentialistic intuitions regarding natural kinds with his famous Twin Earth thought experiment (PUTNAM 1973; 1975).

Philosophers of the special sciences soon began to criticize the traditional view of natural kinds (for a review of the different criticisms, see BIRD and TOBIN 2015). For the most interesting scientific cases, a definition in terms of necessary and sufficient conditions was hard to achieve, and micro-structuralism, as the thesis that kinds can be individuated solely in terms of their microstructural properties, was also put into question. Chemical species and substances and biological species were essentialist philosophers' preferred examples for a long time. Yet, philosophers of chemistry have argued that chemical species and substances are not correctly captured by the essentialist view of natural kinds (NEEDHAM 2000; 2011; VAN BRAKEL 2005; RUTHENBERG 2012). Quite the same happened in the philosophy of biology. Several authors concerned with biological species also turned against the traditional view and argued that an essentialist conception of species is precluded by the "population thinking" that characterizes modern evolutionary biology. The upshot seemed to be that *any* essentialist view of natural kinds was going to be unsuited for biological kinds (MAYR 1966; 1976; see also SOBER 1980). Objections against the traditional essentialist view of natural kinds were not only motivated by a general rejection of natural kind talk in biology altogether (such as the view that species are spatio-temporally extended individuals rather than kinds, cf. GHISELIN 1974 and HULL 1978). Rather, this traditional conception was contested even by advocates of the natural kindhood of biological species (cf. references below).

A different realist view of natural kinds was put forward by Richard Boyd (1991; 1999) and further developed by others (for example MILLIKAN 1999; GRIFFITHS 1999; WILSON, BARKER and BRIGANDT 2007; SLATER 2015), this is, the homeostatic property cluster theory (HPC, hereafter). This view of natural kinds shares an important characteristic with the traditional one, apart from the realist stance. From the point of view of their role in scientific practice, HPC-kinds and traditional kinds stand on an equal footing. However, there are important ways in which HPC theory introduces novel ways of thinking about natural kinds. First, in HPC theory, similarity is nonperfect. Members of a HPC-kind do share some properties, but there is no property or set of properties that can be said to be necessary and sufficient for kind membership. The *cluster* of properties is conceived in an open-ended manner. Thus, HPC-kinds are not definable in terms of sets of necessary and sufficient conditions, as in the Kripke/Putnam view.

But members of a natural kind in the HPC sense are not just imperfectly similar entities. If this was the case, the inductive power of natural kinds would be something of a miracle. HPC-kinds are projectible categories in virtue of sharing homeostatic mechanisms. These mechanisms are causally responsible for the co-occurrence (clustering) of properties and grant their non-accidentalness: entities are not similar *just because*; they are similar by virtue of common mechanisms that tend to maintain certain properties together, or because the instantiation of some properties favor the instantiation of others. In sum, HPC-kinds are defined and individuated by (i) the property cluster and (ii) the set of underlying homeostatic mechanisms responsible for the former.

Another salient feature of HPC theory is that it is compatible with change. Both the property cluster and the set of underlying homeostatic mechanisms may change along with changes in environmental contexts, so HPC-kinds are not immutable. Such a tolerance to change opens up the possibility of kinds with vague boundaries, which makes the theory especially attractive to account for the changing nature of biological phenomena.

There is a widespread understanding that HPC theory (or other philosophical views with the main characteristics of HPC) is better prepared to deal with classification practices in biological sciences, as shown by the multiple applications of it in the philosophy of biology literature. Apart from species (BOYD 1988; 1989; 1991; 1999A; 1999B; KELLER, BOYD & WHEELER 2003; WILSON 2005; WILSON, BARKER and BRIGANDT 2007), other biological categories argued to constitute cluster kinds include higher taxa (RIEPEL 2005B), cell types (WILSON, BARKER and BRIGANDT 2007; SLATER 2013), ecosystems (SLATER 2018), life (DIÉGUEZ 2013; FERREIRA RUIZ and UMEREZ 2018), homologies (RIEPEL 2005A; BRIGANDT 2002), developmental modules (RIEPEL 2005A), and genes (WILSON, BARKER and BRIGANDT 2007). Yet some philosophers have criticized HPC (and/or related views). In the next section, I will consider two objections against cluster kinds and suggest how they can be countered.

### 3. HOW NOT TO RESIST THE NATURAL KIND TALK IN BIOLOGY

The merits of the cluster notions are generally acknowledged, but for some there is a high price to pay for such merits. Here, I will address two objections: *explanatory limitation* and *membership determination*. Countering these objections shows at least two ways in which one should *not* resist the natural kind talk in biology.

### 3.1 *The Explanatory Limitation Objection*

In analyzing different conceptualizations of life, Mark Bedau (2014; 2007) recognizes that a property-cluster approach to living systems neatly explains the existence of borderline cases and why life comes in degrees, difficulties faced by those attempting to define life. However, he also states that cluster views are unsatisfying insofar as they fall short of some explanatory potential that they should bear. Bedau finds the “weakness” of the cluster approaches in the fact that “explanations end too quickly”. He writes:

The view does not explain why certain properties in the cluster are significant and other properties are not (...). The cluster conception cannot explain why life is characterized by one cluster of properties rather than another. The dimensions of the cluster of properties associated with living organisms is simply accepted as a contingent empirical fact. Those who expect *some further explanation* for life’s characteristic properties will find the cluster conception of life unsatisfying (BEDAU 2014, 18, emphasis added)

The objection sounds seriously challenging at first glance, but I want to show that it conflates two different questions and misses the important one. I suggest that the question “Why is a kind characterized by certain properties rather than by others?”, while extremely important to other contexts, is not relevant for evaluating the success or failure of natural kind theories, either essentialistic or otherwise. For an essentialist, since a kind is characterized by its essential properties, and once we’ve found the essence of a kind, it makes little sense to ask further “why is *this* rather than *that* the essence of this kind?” By definition, it could not have been differently (otherwise, this would not really count as an essence). Or, perhaps worse, the only possible answer would be circular: “because that’s its essence”. From the anti-essentialist perspective, the fact that a kind is characterized by certain properties rather than by others (which could be rephrased as the fact that one cluster of properties is maintained by certain homeostatic mechanisms rather than another), while not a matter of definition, could only be regarded either as a brute fact of the universe or as a matter of divine intention. This might sound uncomfortably problematic, but we do seem to accept this type of situation. For example, we would rarely question why the speed of light is 299,792,458 meters per second rather than a different one (LANGE 2002, ch. 4.3). So, it might be the case that a cluster approach cannot offer an answer to this type of questions, but neither can other types of theories, and this is probably unachievable. Bedau’s question shouldn’t be used to evaluate natural kind theories because if it were, no theory would pass the test. This is a situation that could lead us to reject essentialistic and anti-essentialistic views of natural kinds altogether, or could point to something

else. In fact, we might as well wonder whether this was *the* pertinent question—again, from the point of view of a philosophical theory of natural kinds—.

I contend that the crucial question from the point of view of the philosophical problem of natural kinds, and how are natural kind theories evaluated, is *how we distinguish natural kinds from non-natural kinds*. Whereas natural kind theories do not aim at answering Bedau's question, they do offer criteria to answer this one. It is interesting to notice how can both be easily conflated under an essentialist viewpoint: natural kinds are distinguished from non-natural kinds because objects belonging to a given natural kind share essential properties, and those are precisely the properties that characterize the kind. But what is even more interesting is that both questions seem to collapse in Bedau's objection too. Recall he suggests that the cluster approach is flawed because it fails to answer the question as to why a kind is characterized by certain properties rather than by others. If we agree that the key question a theory of natural kinds must pursue is that of how to distinguish natural kinds from non-natural kinds, then, one would also see that his expectation that a cluster approach to natural kinds should, in order to succeed as such (i.e., in order to be able to account for the second), *also* answer the first question, entails a conflation of the two questions<sup>1</sup>.

Of course, I do not mean to deny that there must be an explanation for the fact that, among the multiple properties that characterize living things, some are more common than others. Rather, my point is that this question cannot be properly addressed from the philosophical point of view of identifying natural kinds, where the problem is how to distinguish natural classifications from non-natural ones. In fact, the first question is independent of this latter issue and calls for biological theories, models, generalizations, and hypotheses (SOBER 1992, 763). If we could expect the kind concept of living systems to answer questions such as “why do most living things reproduce?”, or “why do replicating entities usually undergo evolution?” by itself, then what should we need biological science for? Biology must explain the occurrence and co-occurrence of properties of living things, but this is too much to ask of a natural kind concept.

### *3.2 The Membership Determination Objection*

The second objection I consider here has to do not with the explanatory scope of cluster theories when already facing a classification (this is, provided the sorting has been done), but with the sorting itself. More specifically, it

<sup>1</sup> Not in the sense of the failure to distinguish them as distinct questions, but to pick out *that* which expresses the aim of a NK theory.

states that HPC theory (and, likely, any other cluster view) fails to offer solid criteria to fix kind membership. This is due to the fact that both the property family and the underlying homeostatic mechanisms are defined in an open-ended manner, so that both may in principle change (REYDON 2009). Thomas Reydon claims that “HPC theory, then, can only account for kinds whose extensions have already been fixed independently by other means” (2009, 729).

First, Boyd was not unaware of this, and he rather acknowledged that determining membership can be difficult in some cases. However, he didn't see this “necessary indeterminacy”, as he called it (see BOYD 1991, 141-2), as a serious problem for his HPC theory. Other cluster views do not take this to be a problem either (see SLATER 2015). Then, from what point of view is this really a problem?

It will certainly be a serious difficulty for those who believe that a theory of natural kinds must fundamentally provide us with criteria to fix membership, an idea that might betray essentialistic hidden assumptions. From an essentialist point of view, fixing membership is a very simple thing to do *in principle*: one only needs to find out whether the item instantiates the kind's essence, *provided* that said essence has been identified. Shortly, for any given object  $x$ , a given kind  $K$  and an essence  $E$  of  $K$ ,  $x \in K \leftrightarrow Ex$ . Faced with the challenge of identifying essences in the first place, determining membership for a given object is arguably the lesser problem. Nonetheless, the very demand that a theory of natural kinds alone provides with infallible membership determination criteria might be itself an essentialist requirement, as essences are meant to play precisely this role.

It might be true that cluster approaches rely on some independent criteria for membership determination, which are far from agreed-upon for the most part in biological classifications. However, the question is, again: is this a reasonable expectation for natural kind theories? Aren't we assuming (even if implicitly) the requirements self-imposed by an essentialist view of natural kinds? From an anti-essentialist point of view, the argument against cluster kind theories cannot be that these fail to offer reliable criteria to determine membership themselves, as this would amount to claiming that non-essentialistic approaches to natural kinds fail for not being essentialistic! And this wouldn't be fair.

I believe that there are two unwarranted assumptions behind this second objection. First, that no classification system can be said to be *natural* if it fails to classify things in a “black or white”, unambiguous, and definitive manner, insofar as this seems to be a consequence of adopting a cluster approach. And we lack a non-essentialistic argument to support this assumption. Secondly, that any approach to natural kinds that does not enclose or entail membership

determination criteria by itself will fail to be a proper theory of natural kinds. Cluster approaches in fact do not set out to provide an algorithmic procedure to determine membership, because they refuse to foreground one or a few properties as essential to biological classifications. Rather, they aim at providing a general framework for identifying natural kinds *as opposed to arbitrary or mind-dependent groups of things*. I see no reason to believe that determining membership in a clear-cut, “yes or no” manner is a requisite to achieving the former goal and therefore, no reason to believe that *both* things are to be found in a theory of natural kinds. Rather, I contend that the aim of any theory of natural kinds is not so much to be able to fix membership, but to be able to distinguish natural kinds from non-natural kinds in the first place. If the greatest aim of a theory of natural kinds (and the minimum we must expect it to provide for) is rather some criteria with which one can distinguish natural kinds from arbitrary groupings of things, and if cluster theories convincingly achieve this goal, then we need not insist on the membership determination worry. Besides, determining membership is to a great extent a scientific task and should not be left for philosophy alone to settle.

#### 4. FINAL REMARKS

The natural kind talk in biology (and elsewhere) can be resisted from very different points of view. For example, it can be resisted from the more general scientific anti-realism. If science, generally, is not in the business of pursuing truth (or approximations to it), it is not in the business of pursuing natural classification systems in particular, either. Yet someone endorsing such a stance would not bother discussing the peculiarities and details of cluster views of natural kind (being these realistic in character). Perhaps more interesting and challenging, if one has at least minimal realistic intuitions, are those objections that do not stem from a generalized rejection of natural kinds, such as the ones I have addressed here. The objections examined in section 3 are only relevant if one is not previously committed to the view that a natural kind theory must offer empirical explanations, or to essentialism, or to a generalized anti-realism regarding scientific classifications. It is for those who are *not* endorsing such stances that I have tried to dismantle the membership determination objection and the explanatory limitation objection. But if this is not the case, then this is not the appropriate way to resist the natural kind talk in biology—and cluster theories of natural kinds should be argued against in a different way—.

By no means are the arguments herein meant to support a generalized cluster view of natural kinds. In fact, it does not follow from my discussion

of the two objections that a cluster approach to natural kinds is the right way to go. It can still be the case that cluster theories face other, more compelling issues than the alleged pitfalls we have examined. Rather than defending cluster views per se, my aim was to undermine particular objections raised against them and argue that if cluster theories can be resisted, it is not on the grounds of such objections.

The ultimate moral to be drawn from the analysis of the objections reminds our very expectations from philosophical theories of natural kinds: what is it that such a philosophical theory must provide? And this question needs to be properly contextualized, this is, by reference to more general philosophical stances (as scientific essentialism, scientific naturalism, or scientific anti-realism). From an anti-essentialist point of view, rather than considering that the “permissiveness” of cluster approaches must be deemed unacceptable, we might as well revise our own expectations regarding natural kinds.

#### *Acknowledgements*

Thanks to Laura Nuño de la Rosa and Cristina Villegas, organizers of the *Eighth Philosophy of Biology and Cognitive Sciences Research Workshop* in Madrid. I also thank María Cerezo and Vanessa Triviño. Funding for this work was provided by Ministerio de Economía y Competitividad de España (grant FFI2013-47849-P, “Contemporary Questions in the Metaphysics of Biological Sciences”) and CONICET Argentina.



## REFERENCES

- BEDAU, M. 2014, "The Nature of Life", S. LUPER (ed.), *The Cambridge Companion to Life and Death*, Cambridge: Cambridge University Press, 13-29.
- BEDAU, M. 2007, "What is life?", S. SARKAR & A. PLUTYNSKI (ed.), *A Companion to the Philosophy of Biology*, New York: Blackwell, 455-71.
- BIRD, A. & TOBIN, E. 2015, "Natural Kinds", E. ZALTA (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition). URL = <http://plato.stanford.edu/archives/spr2015/entries/natural-kinds/>
- BOYD, R. 1999A, "Homeostasis, Species and Higher Taxa", R. WILSON (ed.), *Species: New Interdisciplinary Essays*, Cambridge: The MIT Press, 141-85.
- BOYD, R. 1999B, "Kinds, complexity and multiple realization", *Philosophical Studies* 95: 67-98.
- BOYD, R. 1991, "Realism, Anti-Foundationalism, and the Enthusiasm for Natural Kinds", *Philosophical Studies* 61: 127-48.
- BOYD, R. 1989, "What realism implies and what it does not", *Dialectica* 43, 5-29.
- BOYD, R. 1988, "How to be a moral realist", G. SAYRE-MCCORD (ed.), *Essays on moral realism*, Ithaca & London: Cornell University Press, 181-228.
- BRIGANDT, I. 2002, "Homology and the origin of correspondence", *Biology and Philosophy* 17 (3): 389-407.
- DIÉGUEZ, A. 2013, "Life as a Homeostatic Property Cluster", *Biological Theory* 7 (2): 180-6.
- FERREIRA RUIZ, M. & UMEREZ, J. 2018, "Dealing with the changeable and blurry edges of living things: a modified version of property-cluster kinds", *European Journal for Philosophy of Science* 8 (3): 493-518.
- GHISELIN, M. 1974, "A Radical Solution to the Species Problem", *Systematic Zoology* 23: 536-44.
- GRIFFITHS, P. 1999, "Squaring the Circle: Natural Kinds with Historical Essences", R. A. WILSON (ed.), *Species: New Interdisciplinary Essays*, Cambridge: MIT Press, 209-28.
- HULL, D. 1978, "A Matter of Individuality", *Philosophy of Science* 45: 335-60.
- KELLER, R. A.; BOYD, R. N. & WHEELER, Q. D. 2003, "The illogical basis of phylogenetic nomenclature", *Botanical Review*, 69: 93-110.
- KORNBLITH, H. 1993, *Inductive Inference and Its Natural Ground*, Cambridge, Massachusetts: MIT Press.
- KRIPKE, S. 1980, *Naming and Necessity*, Cambridge: Harvard University Press.
- LANGE, M. 2002, *An Introduction to Philosophy of Physics: Locality, Fields, Energy and Mass*, Oxford: Blackwell Publishing.

- MAYR, E. 1966, *Animal Species and Evolution*, Boston: Harvard University Press.
- MAYR, E. 1976, *Evolution and the Diversity of Life*, Cambridge: Harvard University Press.
- MILLIKAN, R. 1999, "Historical Kinds and the Special Sciences", *Philosophical Studies* 95: 45-65.
- MILLIKAN, R. 2000, *On Clear and Confused Ideas*, Cambridge: Cambridge University Press.
- NEEDHAM, P. 2000, "What is water?", *Analysis* 60: 13-21.
- NEEDHAM, P. 2011, "Microessentialism: What is the argument?", *Noûs* 45: 1-21.
- PUTNAM, H. 1973, "Meaning and Reference", *Journal of Philosophy* 70: 699-711.
- PUTNAM, H. 1975, "The Meaning of 'Meaning'", *Minnesota Studies in the Philosophy of Science* 7: 215-71
- REYDON, T. 2009, "How to Fix Membership: A Problem for HPC Theory and a Solution", *Philosophy of Science* 72: 724-36.
- RIEPEL, O. 2005A, "Modules, kinds, and homology", *Journal of Experimental Zoology (Molecular and Developmental Evolution)* 304B: 18-27.
- RIEPEL, O. 2005B, "Monophyly, paraphyly, and natural kinds", *Biology and Philosophy* 20: 465-87.
- RUTHENBERG, K. 2012, "What is water? Some philosophical considerations", H. P. HAHN, K. CLESS & J. (ed.), *People at the Well: Kinds, Usages and Meanings of Water in a Global Perspective*, Soentgen, Frankfurt: Campus Verlag, 65-78.
- SLATER, M. 2013, "Cell Types as Natural Kinds", *Biological Theory* 7: 170-9.
- SLATER, M. 2015, "Natural Kindness", *British Journal for the Philosophy of Science* 66: 375-411.
- SLATER, M. 2018, "Anchoring in ecosystemic kinds", *Synthese* 195: 1487-1508.
- SOBER, E. 1980, "Evolution, Population Thinking and Essentialism", *Philosophy of Science* 47: 350-83.
- SOBER, E. 1992, "Learning from Functionalism. Prospects for Strong Artificial Life", C. LANGTON, C. TAYLOR, D. FARMER, & S. RASMUSSEN (ed.), *Artificial Life II*, Massachusetts: Addison-Wesley, 749-65.
- VAN BRAKEL, J. 2005, "On the inventors of XYZ", *Foundations of Chemistry* 7: 57-84.
- WILSON, R. 2005, *Genes and the agents of life: The individual in the fragile sciences—biology*. Cambridge: Cambridge University Press.
- WILSON, R.; BARKER, M. & BRIGANDT, I. 2007, "When Traditional Essentialism Fails: Biological Natural Kinds", *Philosophical Topics* 35 (1-2): 189-215.

JAVIER SUÁREZ & VANESSA TRIVIÑO\*  
*LOGOS, Universitat de Barcelona –Universidad Rey Juan Carlos I de Madrid*

## A metaphysical approach to holobiont individuality: Holobionts as emergent individuals

Received: 6-7-2017 / Accepted: 14-3-2019

**Abstract:** Holobionts are symbiotic assemblages composed by a host plus its microbiome. The status of holobionts as individuals has recently been a subject of continuous controversy, which has given rise to two main positions: on the one hand, holobiont advocates argue that holobionts are biological individuals; on the other, holobiont detractors argue that they are just mere chimeras or ecological communities, but not individuals. Both parties in the dispute develop their arguments from the framework of the philosophy of biology, in terms of what it takes for a “conglomerate” to be considered an interesting individual from a biological point of view. However, the debates about holobiont individuality have important ontological implications that have remained vaguely explored from a metaphysical framework. The purpose of this paper is to cover that gap by presenting a metaphysical approach to holobionts individuality. Drawing upon a conception of natural selection that puts the focus on the transgenerational recurrence of the traits and that supports the thesis that holobionts are units of selection, we argue that holobionts bear emergent traits and exert downward powers over the entities that compose them. In this vein, we argue, a reasonable argument can be made for conceiving holobionts as emergent biological individuals.

**Keywords:** emergence, holobiont, symbiosis, microbiome, biological individuality, trait-recurrence.

\* E-mails: javier.suarez@ub.edu, vanesa.trivino@urjc.es

LIFE ON EARTH DOES NOT WALK ALONE. As it happens in the famous song, 21st century biology has taught us that macrobes (animal and plants)<sup>1</sup> do not walk alone either: they are given in a vast ocean of microbes that find, in different parts of their body surfaces, a comfortable site where they can grow and reproduce. Furthermore, macroorganisms need some of their microorganisms to properly develop and to adequately realize some of their basic functions, including physiological processes like metabolism or immunological defense (DETHLEFSEN *et al.* 2007; GILBERT *et al.* 2012; BRUCKER & BORDENSTEIN 2013; MCFALL-NGAI *et al.* 2013; HILL *et al.* 2016; CHIU & EBERL 2016). The observation of these facts has led some biologists to formulate the so-called *holobiont thesis*. According to this thesis, there is a new level of biological individuality that would encompass the entity formed as a consequence of the association of the macrobe host with all the symbiotic microorganisms that constitute its microbiome, including *Archaea*, bacteria, fungi, and viruses.<sup>2</sup> This new level of individuality is called “holobiont”, and the sum of all the genetic information in a holobiont is called “hologenome” (ROSENBERG *et al.* 2007; see SUÁREZ 2018, for a review of the evolution of the holobiont concept).

Most of the recent debates about holobionts are oriented to either justify or reject their status as biological individuals. On the one hand, advocates of the holobiont thesis argue that the high level of interdependence that exists between macrobes and their symbiotic microorganisms calls for a redefinition of biological individuality. Under the new redefinition, holobionts would be biological individuals in an anatomical, immunological, developmental, physiological, and even in an evolutionary form, constituting a unit of selection in evolution (ZILBER-ROSENBERG & ROSENBERG 2008; DUPRÉ & O’MALLEY 2009; GILBERT *et al.* 2012; 2017; ROSENBERG & ZILBER-ROSENBERG 2014; 2016; DÍAZ 2015; BORDENSTEIN & THEIS 2016; THEIS *et al.* 2016; LLOYD 2017B; ROUGHGARDEN *et al.* 2018).<sup>3</sup>

<sup>1</sup> Throughout the paper, we will refer to animals and plants as “macrobes”/“macroorganisms”, following a famous notation introduced by O’Malley & Dupré (2007). Furthermore, when we refer to the “macrobe host”, we mean the monogenomic individual that derives from a zygote (DUPRÉ 2010; 2012).

<sup>2</sup> “Microbiome” refers to an “entire habitat, including the microorganisms (bacteria, archaea, lower and higher eukaryotes, and viruses), their genomes (i.e., genes), and the surrounding environmental conditions” (MARCHESI & RAVEL 2015, 1), whereas “microbiota” refers to the collection of all the microorganisms (bacteria, fungi, viruses, etc.) that interact in a such environment. In this paper, the host is taken as the environment where those microorganisms interact (in this sense, there is no microbiota without a host where the microbiota grows and reproduces), and “microbiome” and “microbiota” will be used interchangeably. See Lederberg (2000) and Lederberg & McCray (2001) for the first uses of the term.

<sup>3</sup> We cannot provide an exhaustive account of all the different dimensions of biological individuality in this paper. Those readers that are interested can check Clarke (2010), Pradeu (2016A; 2016B), DiFrisco (2017), Lydgard & Nyhart (2017). For our purposes in this paper, we will only concentrate on the evolutionary dimension of biological individuality.

On the other hand, detractors of holobiont individuality claim that the current empirical evidence is not sufficient to claim that holobionts are functionally integrated entities and therefore, any claim about their status as biological individuals is ungrounded. It is important to highlight that detractors of holobionts individuality do not deny that some members of the microbiota might be sufficiently integrated with the host so as to constitute a unique biological individual (e.g. mitochondria in eukaryotes, *Buchnera aphidicola* in aphids, *Wolbachia pipientis* in some insects, etc.). What they deny is that a host plus *all* its symbiotic microbiota act together as a unique biological individual, and find more accurate to talk about independent individuals that live together in association (BOOTH 2014; MORAN & SLOAN 2015; GODFREY-SMITH 2015; CHIU & EBERL 2016; DOUGLAS & WERREN 2016; HURST 2017).

This debate concerning the status of holobionts (i.e. whether they are biological individuals or not) has been mainly approached from a conceptual and an epistemological framework that is proper of the philosophy of biology (see TREVIÑO, unpublished manuscript). Yet, we consider that it has metaphysical consequences for our understanding of mereology, i.e. part-whole relations: which are the metaphysical grounds that move holobiont advocates to argue that they are individuals? How does part-whole relation work in the case of holobionts? This topic, notwithstanding, has never been approached from a purely metaphysical point of view. The purpose of this paper is to cover this gap by elaborating a metaphysical account of the holobiont as an *emergent level of biological individuality*. Drawing upon a conception of natural selection that only requires trait-recurrence among units of selection, we argue that holobionts exhibit metaphysically emergent properties that suggest their consideration as emergent individuals.

Our agenda will be as follows: in §1 we provide an introduction to the philosophical concept of emergence, mainly inspired by Wilson's recent categorization of ontological emergence (2016). Later, in §2, we present the trait-recurrence account of the units of selection, apply it to the holobiont, and explain what this account implies about the status of the holobiont as a biological individual. In §3, we take a metaphysical step by arguing that the trait-recurrence conception of the holobiont entails that holobionts possess properties that allow them to downwardly affect the different entities that compose it (host and microbes of its microbiota). Finally, we conclude the paper by defending that the arguments presented in §1-§3 suggest that, from a metaphysical perspective, holobionts should be conceived as emergent individuals (§4).

## I. ONTOLOGICAL EMERGENCE: WEAK AND STRONG

The idea of emergence was conceptually developed during the late 19th and the early 20th century due to the work of the British Emergentists (McLAUGHLIN 1992, 1997; KIM 2006B). Until the second decade of the 20th century, emergentism was defended as a middle position between reductionism and vitalism. The British Emergentists defended that physical systems, when reaching a certain degree of complexity, can manifest new higher-level properties that are not possessed by the lower-level components of the system nor are reducible to them. Some properties considered as emergent were: life, mind, or chemical bonding.

Due to new explanations in physics accounting for some properties of chemical and biological systems so far considered to be irreducible, as well as the creation and awareness of complex non-linear systems (e.g. turbulent fluids), emergentism disappeared from the mainstream philosophy between the 1930s and 1970s (McLAUGHLIN 1992; WILSON 2013). However, during the last decades of the 20th century, emergentism has reappeared in scientific and philosophical contexts due to the collapse of the reductionistic positions widely defended by the logical positivists (KIM 1989; 1999; 2006B) and the widespread development of the so-called sciences of complexity (GOLDSTEIN 1999; PEPPER & GOLDSTEIN 2004; WITHERINGTON 2011).

Current debates on emergence are oriented towards a conceptual clarification of the basic ideas associated with it, namely dependence and autonomy (VAN GUILICK 2001; CLAYTON 2006; BARNES 2012). However, there is not a unified account of what emergence is yet (KIM 2006B; WILSON 2016), which is problematic. It is not only that the variety of definitions is useless for clarifying the structure of natural reality (WILSON 2016), but also that the philosophical characterizations of emergence are, in some cases, so complex and abstract that it seems really difficult to conceive of an emergent property in the world (McDONALD & McDONALD 2009). However, in different scientific explanations, especially those regarding complex systems, the concept is widely used, and emergent properties are widely recognized. This fact seems to illustrate that there is an incompatibility between how the concept is defined by metaphysicians and how it is used and conceived by philosophers of science. In philosophy of biology, for instance, the abstract characterization of emergent properties in metaphysics is considered to be problematic, insofar as it is unable to make sense of the emergent character of some biological properties, such as the amount of nectar stored in a hive (MITCHELL 2012).

Due to the difficulties in providing an accurate definition of emergence that properly makes sense of its creed: dependence and autonomy, differ-

ent criteria have been proposed to recognize emergent properties. For some authors, emergent properties are unpredictable, irreducible, and have causal powers (KIM 1999). Others have offered a more extensive list of criteria for a property to be emergent, namely holism, novelty, qualitative difference, not given at lower levels, different laws, and interaction (HUMPHREYS 1997). These criteria, as Paul Humphreys makes explicit (HUMPHREYS 1997), are sufficient but not necessary to talk of emergence: they do not need to be met by all emergent properties, nor do all emergent properties need to share the same properties. This implies that different accounts of emergence might be given depending on the criteria one considers emergent properties need to meet (McDONALD & McDONALD 2009). Despite the extant definitions of emergence that are offered, it has been claimed that all of them can be grouped into a twofold categorization, namely *weak ontological emergence* and *strong ontological emergence* (WILSON 2016). Ontological emergence refers to a kind of higher-level properties that are given in the world and that are characterized for being *causally autonomous*, in the sense that the causal power they possess cannot be reduced to the causal power of the lower-level properties of the system upon which they depend (KIM 2006B; JESPER 2006; McDONALD & McDONALD 2009; WILSON 2016). If the higher-level property of a system is not causally autonomous, then there is epiphenomenalism but not emergence.

The notion of causal power used to characterize emergent properties is, notwithstanding, a problematic one. It can be differently conceived depending on the ontological commitments one might have regarding properties. Here, we will follow Wilson's neutral account with respect to causal powers, according to which having a causal power means that the fact that the property is instantiated entails that the entity that bears it will have the power to, in appropriate circumstances, cause some effects, or intervene in the causation of some effects (WILSON 2002; 2016). Furthermore, when it is applied to emergent higher-level properties, it is considered that this causal power needs to be *downwardly exerted* (O'CONNOR 1994, 97-8). Since the 1970s, downward causation is recognized as a central component of emergentism. In philosophy of biology, Donald Campbell suggested that the lower level entities of a system can behave in ways that they would not do, due to the constraints created by the higher level organization that they constitute (CAMPBELL 1974). In this case, we also accept downward causation as one of the hallmarks of emergent properties, and thus, we will consider this feature when arguing for the presence of emergent properties in holobionts.

According to Wilson, the causal autonomous character of an emergent property might be given in two different forms, namely (1) because the



higher-level property has a *new* causal power that is different from the causal powers of the lower-level properties of the system in which the property is instantiated (*strong ontological emergence*)<sup>4</sup>; and (2) because the higher-level property has a mere *proper subset* of the causal powers that the lower-level properties of the system in which it is instantiated have (*weak ontological emergence*).<sup>5</sup> For strong ontological emergence, there is causal autonomy insofar as the emergent property incorporates new causal powers in the world. In cases of weak ontological emergence, conversely, the emergent property does not incorporate new causal powers. Yet, as Wilson claims, the weak emergent property still has a different and autonomous *causal power profile* since it only possesses a *proper subset* of the causal powers of the lower-level properties it depends on (WILSON 2016, 362).<sup>6</sup>

The characterization of emergence is generally attributed to properties. There are properties that emergent or not in a given system. However, in metaphysics, it is accepted that entities might be emergent in a derivational sense from the emergence of properties (BEDAU 2002; WILSON 2016). This is so since, in order for a property to be emergent, the system in which it is instantiated has to reach a certain degree of complexity, such that it is possible to characterize that system as an emergent one. In this regard, when a particular entity is characterized for having an emergent property, then it is considered to be an emergent entity as well. We accept this metaphysical claim, and thus, we consider that if holobionts possess emergent traits, then they are emergent entities.

<sup>4</sup> There is a wide debate in metaphysics concerning whether there are strong ontological emergent properties. An illustration of properties considered to be of this kind are consciousness (CHALMERS 1996), fitness (TREVIÑO, unpublished manuscript), and some properties of quantum mechanics (SILBERSTEIN & MCGEEVER 1999).

<sup>5</sup> An example of weak ontological emergence can be offered in terms of the determinable/determinate relation, such as the relation between colors and shades. Consider that train drivers are said to stop the train when the light of the traffic light is white. 'White' is a determinable that might be multiply determined by different determinates (i.e. the different shades of white) such as 'whitesmoke' or 'snow-white'. Regardless of whether the white of the traffic lights was determined by whitesmoke or snow-white, train drivers must stop the train if they saw the light. In this regard, the different determinates of white share the causal power of the determinable white: advising train drivers they must stop the train. However, they might also have more causal powers besides this one (snow-white might also have the causal power for polar bears to avoid predators, for instance), and they might differ precisely with regard to this other powers. The determinable white, therefore, only inherits a proper subset of the causal power of its determinates, which shows that its causal power profile is different from that of its determinates and that it is not reducible to them.

<sup>6</sup> The causal autonomy Wilson attributes to weak emergent properties on the basis of their different causal power profiles (see also WILSON 1999; SHOEMAKER 2000) might be questionable. In fact, we do not agree with this way of conceiving causal autonomy since it seems to imply reduction. Yet, questioning this form of emergentism goes beyond the main task of this paper.



In §3, we will illustrate that holobionts possess some traits that are causally autonomous insofar as they attribute their bearers a *new* causal power that is not present in the lower-level properties of the parts that compose the holobiont, and that it is also downwardly exerted. Yet, before doing this, we will introduce the conceptual framework from which we conceive units of selection, and thus the role of holobionts as a unit of selection, and explain what this account implies about the status of the holobiont as a biological individual (§2).

## 2. TRAIT-RECURRENCE AS THE KIND OF STABILITY THAT MATTERS IN A HOLOBIONT

One of the most problematic claims about the biological individuality of holobionts comes from the “hologenome concept of evolution” (HCE henceforth). HCE is the thesis that holobionts, with their hologenomes, are a unit of selection in evolution.<sup>7</sup> For an individual to be a unit of selection it has to live in a population of individuals where: (a) different individuals exhibit different phenotypes; (b) different phenotypes are associated with different degrees of fitness; and (c) there is transgenerational inheritance, i.e. the offspring of those individuals that bear a trait *T* will be more likely to bear *T* than the offspring of those individuals that do not bear it. If holobionts are units of selection, as defenders of HCE claim, they have to demonstrate that holobionts satisfy properties (a)-(c).

Most of our current empirical evidence suggests that holobionts satisfy properties (a) and (b). However, the notion that holobionts are units of selection in evolution has been mainly challenged on the basis that there are no proper relations of inheritance across different holobiont generations, and therefore premise (c) is false (MORAN & SLOAN 2015; DOUGLAS & WERREN 2016; HURST 2017). Their argument is grounded on the observation that the species that compose the microbiome of a holobiont do not remain intergenerationally constant, being replaced by different species across different generations of the host. Therefore, they argue, holobiont lineages are not coherent enough to be considered units of selection. Yet, this problem can be addressed if the question about species composition, as well as species transgenerational transmission, is substituted by the notion of transgenera-

<sup>7</sup> Not all the accounts that characterise holobionts as units of selection make a distinction between holobionts and hologenomes. Some authors claim that the holobiont is the interactor, and are mute with respect to the importance of hologenomes, whereas others simply deny the role of hologenomes as replicators. For a summary, see SUÁREZ (2018).

tional trait-recurrence, which is the key element for discussions about the units of selection (SUÁREZ, unpublished manuscript).<sup>8</sup>

In such view of the units of selection, it becomes possible to understand the relations of inheritance across successive generations of holobionts throughout *trait-recurrence*. The main idea would be the following

*A holobiont in generation n+1 will be a coherent unit of selection only if the traits  $T_1, T_2, \dots, T_n$  that appear in the holobiont n+1 are statistically correlated to the traits  $T_1, T_2, \dots, T_n$  that appeared in the holobiont n.*<sup>9</sup>

This account of the units of selection can be characterized by four features. Firstly, the existence of transgenerational trait recurrence among holobionts is dependent upon host-generations (i.e. the host and its offspring are the units taken as referent to count generations of holobionts). As it is generally argued by most holobiont advocates, the life cycle of the holobiont is as long as the life cycle of the host (e.g. GILBERT 2018, 299), thus new generations of holobionts are recognized insofar as new generations of the host are recognized. Secondly, the criterion is theoretically formulated to recognize potential consistent relations of inheritance among holobionts. That is, to recognize possible parent-offspring relations among multispecies consortia<sup>10</sup> that would support the claim that holobionts are units of selection (ZILBER-ROSENBERG & ROSENBERG 2008; DUPRÉ & O'MALLEY 2009; ROSENBERG & ZILBER-ROSENBERG 2014, 2016; BORDENSTEIN & THEIS *et al.* 2016; THEIS *et al.* 2016; GILBERT *et al.* 2017; LLOYD 2017B; ROUGHGARDEN *et al.* 2018). Thirdly, and contrary to other approaches to holobionts that conceive inheritance in terms of species transgenerational transmission (e.g. MORAN & SLOAN 2015; GODFREY-SMITH 2015; STENCEL 2016; DOUGLAS & WERREN 2016; HURST 2017; STENCEL &

<sup>8</sup> In agreement with most recent treatments of units of selection, we assume that an entity in a population is a unit of selection in evolution if it fulfils Lewontin's criteria, i.e. if the entity exhibits heritable phenotypic variance that affects its fitness (LEWONTIN 1970; 1985; OKASHA 2006; GODFREY-SMITH 2009). Contrary to other authors, we do not distinguish between the question of units of selection as interactors versus the question of units of selections as reproducers (e.g. LLOYD 2017A).

<sup>9</sup> One reviewer has rightly pointed out that the model of trait recurrence is not exclusive for holobionts, but for every unit of selection. We agree with her that the account should work for every unit of selection. However, because in this paper we focus only on the case of the holobiont, we only restrict our comments to the consequence that the account would have for holobionts.

<sup>10</sup> Not all multispecies consortia are holobionts (e.g. biofilms, ecosystems), and our definition is formulated exclusively for holobionts, insofar as it assumes the existence of the host as the criterion that grants intergenerational identity, a circumstance that does not need to occur in all multispecies consortia.

WLOCH-SALAMON 2018), in this account, inheritance is reformulated in terms of reappearance of traits, i.e. reappearance of structures which, *ceteris paribus*, increase the fitness of the entities that bear them (TRIVIÑO & NUÑO DE LA ROSA 2016). Finally, the existence of non-species related trait-recurrence in holobionts (and thus, the validity of trait-recurrence as a conceptual criterion for defining holobionts as units of selection) is empirically supported by recent discoveries that suggest that some traits in macrobes appear as a consequence of their interaction with a functionally-equivalent microbiome, which, despite being taxonomically different for hosts of the same species, can still bring about the same metabolic traits through their interactions with the hosts (e.g. BURKE *et al.* 2011; TAXIS *et al.* 2015; HESTER *et al.* 2016; LOUCA *et al.* 2016; LEMENCEAU *et al.* 2017).

According to the criterion of trait-recurrence, then, holobionts are biological individuals in virtue of their ability to reconstruct their traits transgenerationally by recruiting the members of their microbiota. One worry might arise at this point: why is the holobiont, and not the host, the entity that gets its traits reconstructed transgenerationally? The answer: because the recruitment of the members of the microbiota is not necessarily determined by the host, but co-determined by the interactions between the host and the microbiota (e.g. GILBERT & CHIU 2015; CHIU & EBERL 2016). There are three different channels to guarantee the acquisition of the microbiome: first, vertical transmission during conception (e.g. mitochondria in the eukaryotic cell; primary symbionts in many insects, including *B. aphidicola* in aphids) (BRIGHT & BULGHERESI 2010); second, direct transmission during birth, or during weaning (FUNKHAUSER & BORDENSTEIN 2013); and third, environmental acquisition, either through family/social interaction or through diet. The first two forms of acquisition do not put any pressure on the traditional definition of inheritance, as it is expected that they will happen by species-transmission. The third one, however, can only be conceptually understood by appealing to the notion of the units of selection just introduced, as in many cases of environmental acquisition of the microbiome there is no species constancy (i.e. the holobiont in generation  $n+1$  does not environmentally acquire the same species of microbes that the holobiont in generation  $n$  has acquired), despite the existence of *transgenerational trait-recurrence*.

Assuming: first, that those holobionts that acquire microorganisms that bring about fitter traits will (*ceteris paribus*) survive longer and reproduce more<sup>11</sup>; and second, that there are mechanisms that guarantee the survival of

<sup>11</sup> That is to say, assuming that no other evolutionary factors (or forces) are putting pressure on the evolution of the population; i.e., assuming that natural selection is the only factor determining how the population evolves.

the microorganisms outside the holobiont (e.g. BROWNE *et al.* 2017), as well as the existence of a higher degree of horizontal gene transfer among the members of a microbiome (e.g. LIU *et al.* 2012; LERNER *et al.* 2017). Then, there are three mechanisms that guarantee that environmental transmission of traits will be preserved across generations of holobionts in spite of the lack of the perfect vertical transmission of the species of the microbiome. These mechanisms are: (1) the genetics of the host, such that a host in generation  $n+1$  will be more genetically predisposed to acquire functionally similar microorganisms than the host in generation  $n$ ; (2) the immunological characteristics of the host, such that a host in generation  $n+1$  will be more immunologically predisposed to acquire functionally similar microorganisms than the host in generation  $n$ ; (3) the contact of the host with an environment where functionally similar microorganisms can easily survive and therefore tend to be present. These three mechanisms together increase the likelihood that a holobiont in generation  $n+1$  will bear traits that are more similar to a holobiont in generation  $n$  which is its progenitor, than to a random member of the population where it lives. This, therefore, suggests that holobionts can be conceived as biological individuals *qua* units of selection, in so far as the condition of inheritance is now satisfied by appealing to trait-recurrence.

### 3. FROM TRAIT-RECURRENCE TO DOWNWARD CAUSATION

In the previous sections (§1 and §2) we introduced two separated nodes, the first one regarding the concept of emergence, and the second one about how to conceive holobionts, so that the notion that they are biological individuals *qua* units of selection makes sense. Here we connect the two nodes to argue that the conception of holobionts as generations of stabilized traits entails the metaphysical consequence that holobionts are emergent entities. In order to do so, we will illustrate that some of the traits that characterize holobionts have a new causal power that is downwardly manifested.

As we argued in §2, transgenerational transmission among holobionts can be understood in terms of trait-recurrence: the likelihood that a trait  $T$  that appears in a holobiont in generation  $n$  will reappear in a holobiont in generation  $n+1$  is higher than the likelihood that  $T$  appears in a random member of the population  $P$ . Furthermore, as we argued, if  $T$  increases the fitness of the holobiont, then *ceteris paribus*, it will be driven towards fixation in  $P$ . Because this can be so, then natural selection at the level of the holobiont is conceptually possible, as well as empirically verified in some cases (e.g. BRUCKER & BORDENSTEIN 2012; MENDOZA *et al.* 2018). An immediate consequence of the

trait-recurrence model of holobiont selection is that the holobiont will exercise causal powers over the entities that compose it. In other words, because of the transgenerational trait-recurrence and selection at the level of the holobiont, the different entities that compose it will be the action of natural selection sorted. As a consequence, some of them will be eliminated from the environment, others will increase their frequency, and yet others will evolve so that they are able to fulfill the functions that cause the appearance of the traits that increase the fitness of the holobiont and, *ceteris paribus*, will tend to get fixed in *P*. This, we argue, is a form of *downward causation* at the level of the holobiont, since the holobiont determines which elements of the microbiome succeed transgenerationally and which do not, thus conditioning their evolutionary dynamics.

We can illustrate how trait-recurrence entails downward causation by means of an example. Let us take the case of the aphid holobiont (aphid + *Buchnera aphidicola*).<sup>12</sup> As it is well known, some elements of the holobiont's phenotype strongly depend on the interaction between the host and its microbiome. One standard trait that has been empirically studied is thermal tolerance, which is known to depend on the interaction between the host and the microbiome, therefore constituting a standard case of a holobiont trait (see DUNBAR *et al.* 2007). The thermal tolerance of an organism dictates the range of temperatures where the organism can live and reproduce, therefore determining the types of environments where it is viable. As a trait, it confers certain selective advantages in those cases where the range of thermal tolerance is higher, as it means that the organism would be viable in a wider range of environments and thus, *ceteris paribus*, it will live longer and reproduce more. Importantly, hence, those aphid holobionts that have a higher degree of thermal tolerance will live longer and reproduce more than any other average aphid holobiont in the population. This will have two consequences: First, those aphid hosts whose holobiont has a higher degree of thermal tolerance will survive longer and reproduce more than an average member of the population; second, those *B. aphidicola* that are part of the microbiome with a higher degree of thermal tolerance will also survive longer and, more importantly, they will have more opportunity of spreading in the population of holobionts.

The second case, i.e. the spreading of *B. aphidicola* in the holobiont population, is a case of downward causation, where it is the holobiont the entity that causes *B. aphidicola* to spread in the population, therefore conditioning its evolutionary dynamics. To explain why, we need to start supposing that the

<sup>12</sup> We are conscious that the aphid holobiont would also include other elements of the aphid microbiome. Since our paper is not about philosophy of biology, but about metaphysics, we restrict our argument to the case of the aphid + *B. aphidicola*, which as our reviewer has correctly pointed out is not problematic for any of the contenders in the holobiont debate.

holobiont population is only affected by the forces of natural selection, which would totally determine its evolutionary dynamics. We know that this situation is highly ideal and far from what happens in reality, where many evolutionary factors affect populations simultaneously, having sometimes opposing effects that cancel each other. However, this would not affect our metaphysical point: if we can demonstrate that, *ceteris paribus*, there is downward causation, then no matter what evolutionary factors would cancel their effects, there would be downward causation acting on *B. aphidicola* yet.

Now, supposing that some of the holobionts in the population will bear the trait that confers fitness advantages whereas others do not, the population can be divided into two groups: individuals that bear the trait (*A*), and individuals that do not bear the trait (*B*). It is known, by hypothesis, that the individuals of *A* will survive longer and reproduce more than the individuals of *B*. If this is the case, then the *B. aphidicola* that compose the microbiome of the individuals in *A* will also survive longer than the *B. aphidicola* of the individuals in *B*. Since it will survive longer, there will be more opportunity for spreading: the holobionts of *A* will have more physical contact with other holobionts than those in *B*. Because this is so, it is more likely that their *B. aphidicola* will get passed, in the scenario where it is acquired by horizontal transmission: more contacts means more opportunity for getting passed, and more opportunity for getting passed means more likelihood of getting fixed in the population.<sup>13</sup> This is so because of the opportunity of longer contact with those holobionts in *A* (by hypothesis they live longer) and the different opportunities that microorganisms have to survive in the environment.

On the other hand, in the case that the *B. aphidicola* are transmitted vertically, it will also be the case that they will increase their numbers in the population necessarily, as those holobionts that bear the trait that confer fitness advantages will, by hypothesis, reproduce more than those that do not. If this is so, then we would have a case where the selection of these holobionts that bear the trait that confers the fitness advantages determines that those microorganisms that constitute their microbiome are sorted differentially. The causation, thus, works from the higher-level entity to its lower level parts, therefore constituting an example of downward causation.

Two metaphysical consequences about the ontological status of holobionts can be deduced from the previous example. Firstly, we argue that the form of dependence that exists between the traits in a holobiont and the lower-level

<sup>13</sup> The case developed here is supposed to cover every possible hypothetical scenario. It is known that *B. aphidicola* is only transmitted vertically in aphids. But many of the members of the microbiome in different species are transmitted horizontally. And this is the only thing that matters for our conceptual purposes here.

parts that determine the appearance of those traits is a form of strong ontological emergentism (WILSON 2016). This is so for two reasons: first, because those traits appear and are maintained as a consequence of the interaction between the host and the microbiome; in other words, the host without the microbiome or the microbiome without the host do not *cause* those traits to appear, which suggests that they are specific to the holobiont and therefore non reducible to the host (irreducibility of the traits); and second, because those traits are usually multiply-realizable; that is, those traits appear as a consequence of the interaction between the host and a microbiome, but the components of the microbiome that cause those traits to appear might differ among different hosts of the same species, while the same traits still appear (multiple realizability of the traits). Secondly, we argue that the existence of downward causation, insofar as it implies strong ontological emergentism (WILSON 2016), entails the possibility of conceiving holobionts as emergent individuals in virtue of having emergent traits. In this sense, holobionts would be a new level of biological individuality, with their own properties (traits) that have causal consequences on the parts that constitute them and, therefore, a genuine new level in the biological hierarchy.

#### 4. CONCLUDING REMARKS: HOLOBIONTS AS EMERGENT INDIVIDUALS

This paper argued that holobionts can ontologically be conceived as emergent individuals on the basis that they possess properties that allow them to exercise downward causation on the entities that compose them. Drawing upon the trait-recurrence model of holobiont individuality, we argued that a consequence of holobionts selection was the sorting of the entities that constitute the holobiont. In other words, we argued that the holobiont, as a unit of selection, exercises downward causation over the entities that constitute it, causing its differential degrees of evolutionarily success or failure. We further argued that the existence of this downward causation was compatible with strong emergentism and, therefore, with the hypothesis that holobionts are emergent ontological individuals.

#### *Acknowledgments*

We would like to acknowledge Laura Nuño de la Rosa, and Cristina Villegas for inviting us to participate in this special issue, as well as for their careful reading of the first version of the paper. We also want to extend this acknowledgment to our anonymous referee, who had the patience to give us a really valuable feedback that helped us to improve our initial draft. Javier

Suárez would like to formally acknowledge the Spanish Ministry of Education (FFU16/02570) and the Spanish Ministry of Economy and Competitiveness (FFI2016-76799-P) for their economic support. Vanessa Triviño would also like to thank the “Fundación Séneca. Agencia de Ciencia y Tecnología de la Región de Murcia” (19489/PI/14), and the Spanish Ministry of Economy and Competitiveness (FFI2017-87193-P).



## REFERENCES

- BARNES, E. 2012, "Emergence and Fundamentality", *Mind*, 121: 873-901.
- BOOTH, A. 2014, "Symbiosis, selection and individuality", *Biology and Philosophy*, 29: 657-73.
- BORDENSTEIN, S. & THEIS, K. 2016, "Host biology in the light of the microbiome: Ten principles of holobionts and hologenomes". *PLOS Biology*. DOI:10.1371/journal.pbio.1002226.
- BURKE, C.; STEINBERG, P.; RUSCH, D.; KJELLBERG, S. & THOMAS, T. 2011, "Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences*, 108 (34):14288-93.
- CAMPBELL, D. T. 1974, "Downward causation in hierarchically organized biological systems", AYALA, F. J. & DOBZHANSKY, T. (ed.), *Studies in the Philosophy of Biology*, Berkeley/Los Angeles: University of California Press, 179-86.
- CHALMERS, D. 1996, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.
- CHIU, L. & EBERL, G. 2016, "Microorganisms as scaffolds of biological individuality: An eco-immunity account of the holobiont", *Biology and Philosophy*, 31: 819-37.
- CLARKE, E. 2010, "The problem of biological individuality", *Biological Theory*, 5 (4): 312-25.
- CLAYTON, P. 2006, "Conceptual Foundations of Emergence Theory", *The Re-Emergence of Emergence*, Oxford: Oxford University Press.
- DÍAZ, J. S. 2015, "El mecanismo evolutivo de Margulis y los niveles de selección", *Contrastes. Revista Internacional de Filosofía*, XX (1): 7-24.
- DIFRISCO, J. 2017, "Kinds of biological individuals: Sortals, projectability, and selection", *The British Journal for the Philosophy of Science*.
- DOOLITTLE, W. F. & BOOTH, A. 2017, "It's the song not the singer: an exploration of holobiosis and evolutionary theory", *Biology and Philosophy*, 32: 5-24.
- DOUGLAS, A. & WERREN, J. 2016, "Holes in the hologenome: why host-microbe symbioses are not holobionts", *mBio*, 7 (2): e02099-15.
- DUPRÉ, J. 2010, "The polygenomic organism", *The Sociological Review*, 58 (s1): 19-30.
- DUPRÉ, J. 2012, *Processes of Life: Essays in the Philosophy of Biology*, Oxford: Oxford University Press.
- FUNKHOUSER, L. J. & BORDENSTEIN, S. R. 2013, "Mom Knows Best: The Universality of Maternal Microbial Transmission", *PLoS Biology* 11 (8): e1001631.
- GILBERT, S.; SAPP, J. & TAUBER, A. 2012, "A symbiotic view of life: We have never been Individuals", *The Quarterly Review of Biology*, 87 (4): 325-41.

- GILBERT, S.; ROSENBERG, E. & ZILBER-ROSENBERG, I. 2017, "The holobiont with its hologenome is a level of selection in evolution", GISSIS S.; LAMM, E. & SHAVIT, A. (ed.), *Landscapes of collectivity in the life sciences*, London: The MIT Press, 305-24.
- GOLDSTEIN, J. 1999, "Emergence as a Construct: History and Issues", *Emergence*, 1: 49-72.
- HESTER, E.; BAROTT, K.; NULTON, J.; VERMEIJ, M. & ROHWER, F. 2016, "Stable and sporadic symbiotic communities of coral and algal holobionts", *The ISME Journal*, 10: 1157-69.
- HUMPHREYS, P. 1997, "How Properties Emerge", *Philosophy of Science*, 64: 1-17.
- HURST, G. 2017, "Extended genomes: symbiosis and evolution", *Interface Focus*, 7: 20170001.
- JESPER, K. 2006, "The causal exclusion argument", *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 131 (2): 458-85.
- KIM, J. 1989, "The Myth of Non-reductive Materialism", *Proceedings and Addresses of the American Philosophical Association*, 63 (3): 31-47.
- KIM, J. 1999, "Making sense of Emergence", *Philosophical Studies*, 95: 3-36.
- KIM J. 2006B, "Emergence: Core ideas and Issues", *Synthese*, 151: 547-59.
- LEMANCEAU, P.; BLOUIN, M.; MULLER, D. & MOËNNE-LOCCOZ, Y. 2017, "Let the core microbiota be functional", *TRENDS in Plant Science*, 22 (7): 583-95.
- LLOYD, E. 2017A, "Units of selection", E. N. ZALTA (ed.), *The Stanford Encyclopaedia of Philosophy*. <https://plato.stanford.edu/entries/selection-units/>
- LLOYD, E. 2017B, "Holobionts as units of selection: Holobionts as interactors, reproducers, & manifestors of adaptation", GISSIS, S.; LAMM, E. SHAVIT, A. (ed.), *Landscapes of collectivity in the life sciences*, London: MIT Press, 351-67.
- LOUCA, S.; JACQUES, S.; PIRES, A.; LEAL, J.; SRIVASTAVA, D.; PARFREY, L.; FARJALLA, V. & MICHAEL, D. 2016, "High taxonomic variability despite stable functional structure across microbial communities", *Nature Ecology and Evolution*, 1: 15.
- MACDONALD, C. & MACDONALD, G. 2009, Emergence and Downward Causation.
- MCLAUGHLIN, B. 1992, "The Rise and Fall of British Emergentism", BECKERMAN, A.; FLOHR, H. & KIM, J. (ed.), *Emergence or Reduction? Essays on the Prospects of Non-Reductive Physicalism*, Berlin: De Gruyter, 49-93.
- MCLAUGHLIN, B. 1997, "Emergence and Supervenience", *Intellectica*, 2: 25, 25-43.
- MITCHELL, S. 2012, "Emergence: Logical, Functional and Dynamical", *Synthese*, 185 (2): 171-86.
- MORAN, N. & SLOAN, D. 2015, "The Hologenome Concept: Helpful or Hollow?", *PLoS Biol*, 13 (12): e1002311.
- O'CONNOR, T. 1994, "Emergent Properties", *American Philosophical Quarterly*, 31: 91-104.

- O'MALLEY, M. & DUPRÉ, J. 2007, "Size doesn't matter: Towards a more inclusive philosophy of biology", *Biology and Philosophy*, 22: 155-91.
- PEPPER, S. & GOLDSTEIN, J. 2004, *Emergence: Complexity and Organization*. DOI: 10.emerg/10.17357.8163f5b43a8f4de73879e679bb0a5b95
- PRADEU, T. 2016A, "The many faces of biological individuality", *Biology and Philosophy*, 31: 761-73.
- PRADEU, T. 2016B, "Organisms or biological individuals? Combining physiological and evolutionary individuality", *Biology and Philosophy*, 31: 797-817.
- ROSENBERG, E.; KOREN, O.; RESHEF, L.; EFRONY, R. & ZILBER-ROSENBERG, I. 2007, "The role of microorganisms in coral health, disease and evolution", *Nature Reviews: Microbiology*, 5: 355-62.
- ROSENBERG, E. & ZILBER-ROSENBERG, I. 2014, *The Hologenome Concept*, London: Springer.
- ROSENBERG, E. & ZILBER-ROSENBERG, I. 2016, "Microbes drive evolution of animals and plants: the hologenome concept", *mBio*, 7 (2): e01395-15.
- ROUGHGARDEN, J.; GILBERT, S.; ROSENBERG, E.; ZILBER-ROSENBERG, I. & LLOYD, E. 2017, "Holobionts as Units of Selection and a Model of their Population Dynamics and Evolution", *Biological Theory*, 13: 44-65.
- SHOEMAKER, S. 2000, "Realization and Mental Causation", *Proceedings of World Congress in Philosophy*, Cambridge: Philosophy Documentation Center.
- SHROPSHIRE, J. & BORDENSTEIN, S. 2016, "Speciation by symbiosis: the microbiome and behavior", *mBio*, 7 (2): e01785-15.
- SILBERSTEIN, M. & MCGEEVER, J. 1999, "The search for ontological emergence", *The Philosophical Quarterly*, 48: 125, 182-200.
- SKILLINGS, D. 2016, "Holobionts and the ecology of organisms: Multi-species communities or integrated individuals?", *Biology and Philosophy*, 31: 875-92.
- STENCEL, A. 2016, "The relativity of Darwinian populations and the ecology of endosymbiosis", *Biology and Philosophy* 31 (5): 619-37.
- STENCEL, A. & WŁOCH-SALAMON, D. M. 2018, "Some theoretical insights into the hologenome theory of evolution and the role of microbes in speciation", *Theory in Biosciences*, 137 (2): 197-206.
- SUÁREZ, J. 2018, "The importance of symbiosis in philosophy of biology: An analysis of the current debate in biological individuality and its historical roots", *Symbiosis*, 76 (2): 77-96.
- SUÁREZ, J. (unpublished manuscript), "Stability of traits as the kind of stability that matters. Holobionts as units of selection from a multilevel selection framework".
- TAXIS, T.; WOLFF, S.; GREGG, S.; MINTON, N.; ZHANG, C.; DAI, J.; SCHNABEL, R.; TAYLOR, J.; KERLEY, M.; PIRES, J.; LAMBERSON, W. & CONANT, G. 2015, "The

- players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity”, *Nucleic Acids Research*, 43 (20): 9600-12.
- THEIS, K.; DHEILLY, N.; KLASSEN, J.; BRUCKER, R.; BAINES, J.; BOSCH, T.; CRYAN, J.; GILBERT, S.; GOODNIGHT, C.; LLOYD, E.; SAPP, J.; VANDENKOORNHUYSE, P.; ZILBER-ROSENBERG, I.; ROSENBERG, E. & BORDENSTEIN, S. 2016, “Getting the hologenome concept right: an eco- evolutionary framework for hosts and their microbiomes”, *mSystems*, 1 (2): e00028-16.
- TRIVIÑO, V. (unpublished manuscript). “A study of the interactions between metaphysics and biology: towards a metaphysics of biology.”
- TRIVIÑO, V. & NUÑOS DE LA ROSA, L. 2016, “A Causal Dispositional Account of Fitness”, *History and Philosophy of the Life Sciences*, 38 (3): 1-18.
- VAN GULICK, R. 2001, “Reduction, Emergence and other Recent Options on the Mind/Body Problem: a Philosophical Overview”, *Journal of Consciousness Studies*, 8: 9-10, 1-34.
- WILSON, J. 1999, “How Superduper does a Physicalists Supervenience Need to be?”, *The Philosophical Quarterly*, 49: 33-52.
- WILSON, J. 2002, “Causal Powers, Forces and Superdupervenience”, *Grazer Philosophische Studien*, 63: 53-78.
- WILSON, J. 2013, “Nonlinearity and Metaphysical Emergence”, S. MUMFORD & M. TUGBY (ed.), *Metaphysics and Science*, Oxford: Oxford University Press, 201-35.
- WILSON, J. 2016, “Metaphysical emergence: weak and strong”, T. BIGAJ & C. WÜTHRICH (ed.), *Metaphysics in Contemporary Physics*, Boston: Brill Rodopy, 345-402.
- WITHERINGTON, D. 2011, “Taking Emergence Seriously: The Centrality of Circular Causality for Dynamic Systems Approaches to Development”, *Human Development*, 54: 66-92.
- ZILBER-ROSENBERG, I. & ROSENBERG, E. 2008, “Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution”, *FEMS Microbiol Rev*: 723-73.

# *Brúixola filosòfica*



VIRGINIA BALLESTEROS<sup>1</sup>  
*Universitat de València*

## Quines normes? Una aproximació al debat sobre el concepte de malaltia mental

*Which norms? An approach to the debate on the concept of mental illness*

**Resum:** En aquest article realitzem una aproximació al debat sobre el concepte de malaltia mental, des dels anys cinquanta fins al present, incorporant-hi elements de la filosofia i sociologia de la medicina, i atenent a les seues dimensions orgànica, subjectiva i social. Perfilem i avaluem les postures naturalistes i normativistes que han conformat el debat sobre la malaltia mental: des de l'antipsiquiatria a la psiquiatria crítica, passant per un naturalisme i construccionisme social forts.

**Abstract:** In this paper we approach the debate on the concept of mental illness from the 1950s to the present, by incorporating elements from the philosophy and sociology of medicine, considering its bodily, subjective, and social dimensions. We outline and evaluate the naturalist and normativist views that have shaped the debate on mental illness: from antipsychiatry to critical psychiatry, passing through strong naturalism and social constructionism.

**Paraules clau:** malaltia mental, naturalisme, construccionisme, antipsiquiatria, psiquiatria crítica.

**Keywords:** mental illness, naturalism, constructionism, antipsychiatry, critical psychiatry.

<sup>1</sup> Aquesta investigació està finançada per l'Ajuda per a la Contractació de Personal Investigador en Formació (ACIF) de caràcter predoctoral que l'autora percep del Fons Social Europeu (FSE) i la Conselleria d'Educació, Investigació, Cultura i Esport de la Generalitat Valenciana.

SI PENSEM EN EL CONCEPTE de *malaltia mental* es fa patent que tant la *malaltia* com allò *mental* poden ser objecte de grans controvèrsies. Definir la malaltia no és menys complicat que definir allò mental; i, quan tots dos conceptes van de la mà, sorgeixen problemàtiques particulars. El propòsit d'aquest article és el de realitzar una aproximació als debats que han sorgit entorn de la definició i realitat de la malaltia mental des dels anys 60 fins al present. La pregunta que ens guiarà pel nostre recorregut serà la de *quines normes*; de quines normes es desvia la malaltia mental: naturals o socials? Ens aproximarem a postures a un costat i a l'altre de la dicotomia, que podem entendre com a normativistes i naturalistes; sense oblidar-nos d'aquells que veuen la virtut en el terme mitjà i opten per postures híbrides.

Ara com ara, per començar a dibuixar el panorama, podem definir les postures naturalistes com aquelles que consideren que el concepte de malaltia mental no conté qüestions de valor; dient-ho de manera positiva, que únicament capta fets naturals —usualment una alteració o disfunció biològica, en algun sentit—. Per una altra banda, les postures normativistes consideren que la malaltia mental no rastreja cap fet natural, sinó merament relacions socials o qüestions de valor: els malalts mentals ho són en virtut del fet que presenten emocions o comportaments sancionats negativament per la societat i cultura en què es troben. En paraules de Robert Kendell:

La major part del debat explícit sobre el concepte de trastorn mental des de 1960 ha girat al voltant de la qüestió de si les atribucions de trastorn mental descansen en algun fet mèdic incontestable o si són expressió de normes i valors socials (KENDELL 1986, 25)

Lligades a la problemàtica de la definició del concepte de malaltia mental, apareixen a més tota una sèrie de problemes relacionats amb els termes que conformen el seu camp semàntic. S'observarà, per exemple, que l'autor recentment citat, Robert Kendell, no parla de *malaltia mental*, sinó de *trastorn mental*. Encara que cada autor utilitza preferentment un dels dos termes, la veritat és que no hi ha un significat universalment acceptat per a cadascun d'ells: precisament per això escrivim aquest article. No obstant això, hi ha autors que prefereixen el terme *trastorn* per a evitar tot el bagatge que porta associat el terme *malaltia*. En aquest sentit, és possible que *malaltia* ens remeta a *malaltia física*, de manera que podria pensar-se que es parla de la malaltia mental de manera anàloga a la malaltia del cos. El terme *trastorn* no sembla tindre necessàriament associat aquell significat. Així i tot, això és sovint una il·lusió: no són pocs els qui parlen de trastorns mentals compresos de manera física. Per a evitar confusions, posarem especial èmfasi a adoptar la terminologia de cada autor i explicar-ne l'ús.



La segona qüestió que convé assenyalar és que la nostra llengua parteix d'una limitació a l'hora de parlar de la malaltia que no trobem en la llengua en la qual es formula la major part d'aquest debat, l'anglès. Així, en la llengua anglesa tenim tres termes que apunten a tres dimensions del nostre terme *malaltia*: *disease*, *illness* i *sickness*. Aquesta distinció, que fou aplicada notablement dins del camp de la sociologia de la medicina en els anys cinquanta, ha donat forma a bona part del debat posterior i suposa una eina important per a comprendre els diferents vessants de la malaltia, siga física o mental; encara que, com tot producte intel·lectual, també compta amb els seus detractors i no és emprada per tothom (HOFMANN 2002).

Per tal d'introduir aquest punt, podem assenyalar que *disease* fa referència a la dimensió orgànica, fisiològica, de la malaltia; *illness*, a l'experiència subjectiva que el malalt té del seu patiment o la seua afecció; i *sickness*, a la dimensió social de la malaltia, al seu significat i al rol de malalt. Segons la caracterització d'Andrew Twaddle (citat per HOFMANN 2002, 652-3), podem definir les tres dimensions de la malaltia de la manera següent:

*Disease* és un problema de salut que consisteix en un mal funcionament fisiològic que resulta en una reducció real o potencial de les capacitats físiques i/o una reducció de l'esperança de vida.

*Illness* és un estat de salut indesitjable interpretat subjectivament. Consisteix en estats de sentiments subjectius (per exemple, dolor, feblesa), percepcions de l'adequació del seu funcionament corporal i/o sentiments de competència.

*Sickness* és una identitat social. És la mala salut o el problema o problemes de salut d'un individu definit per uns altres amb referència a l'activitat social d'aquest individu.

Clarament, *disease*, *illness* i *sickness* no són dimensions estanques, sinó que es troben en constant interacció: és obvi, per exemple, que el rol social que es conferisca al malalt tindrà un impacte considerable en la seua pròpia experiència subjectiva; i, anant més enllà, també observem que tindrà un impacte en la dimensió orgànica, atès que les estructures socials disponibles per al tractament de la seua malaltia interferiran amb el seu curs.

Encara que no tots els autors elaboren les seues propostes atenent a aquesta tríada i emprant les mateixes definicions, comprendre aquestes tres dimensions ens resulta útil per començar a disposar algunes peces sobre el nostre tauler filosòfic. El primer moviment que aquestes peces ens permeten fer és comprendre que el debat més acarnissat no versa sobre l'existència o inexistència de les dimensions subjectives o socials de la malaltia mental, sinó sobre la realitat i les possibilitats de coneixement de la malaltia mental com a fet natural, és a dir,

com a *disease*. En referència a això s'han produït dos debats importants: un d'ells en les files dels mateixos naturalistes i partidaris de les posicions híbrides, en què es discuteix sobre com ha de comprendre's exactament el significat de *malaltia orgànica* o *trastorn* —com a lesió anatòmica, com a disfunció biològica, etc.— i quin impacte té la seua definició en la malaltia o trastorn mental<sup>2</sup> (BOORSE 1976; KENDELL 1975; SCADDING 1967; SZASZ 1972; WAKEFIELD 1992); i l'altre debat, entre normativistes i naturalistes, que se centra en la possibilitat mateixa que el vessant orgànic de la malaltia capture realment fets naturals o, per contra, remeta merament a fets socials (EISENBERG 1988; ENGELHARDT 1976; SEDGWICK 1982; TIMMERMANS i HAAS 2008; TURNER 1995).

Farem finalment una última i breu observació respecte a un altre terme d'aquest camp semàntic; tot i ser un concepte molt pròxim al de malaltia, no hem entrat en el debat respecte al concepte de *salut*. Sobre això tan sols direm que les aproximacions positives a aquest concepte no solen abundar. De forma positiva, la salut se sol definir com a funcionament òptim; mentre que, de forma negativa, sol definir-se com a absència de malaltia.<sup>3</sup>

Per a concloure aquesta introducció, assenyalem un aspecte de la malaltia mental que anirem subratllant al llarg d'aquest treball i que resulta obvi per poc que es reflexione sobre la qüestió: les enormes implicacions pràctiques que té la seua atribució. Dir d'algú que té una malaltia és una *crida a l'acció*, això és, a la intervenció terapèutica (SPITZER, ENDICOTT i MICOULAUD FRANCHI 2018). L'atribució de malaltia mental implica que alguna cosa no és com *hauria de ser* i, a més, que probablement no es corregirà per si mateixa, sinó que requereix intervenció professional —en aquest cas, d'un psiquiatra—. A més, la malaltia mental també té repercussions importants en el camp de la responsabilitat moral i legal, ja que aquesta n'és un atenuant o eximent, atès que es considera que el malalt té minvades les seues capacitats cognitives i/o volitives. Així mateix, el concepte de malaltia mental també té un gran impacte en la subjectivitat individual, així com en el rol social de la persona que és diagnosticada. És per les seues implicacions pràctiques que és crucial comprendre la naturalesa del concepte de malaltia mental: presentar-lo com a constructe social, i per tant contingent, obri la porta a la seua crítica i revisió; per contra, presentar-lo dins del món natural revist el concepte de la legitimitat pròpia de la necessitat de la naturalesa.

<sup>2</sup> Cal posar en relleu ací que un pot ser naturalista respecte a la malaltia física i normativista respecte a la malaltia mental. És precisament per això que hem inclòs l'autor Thomas Szasz com a naturalista; explicarem amb detall la seua postura més endavant. Encara que usualment es considera com a revisionistes les postures normativistes, la veritat és que el mateix naturalisme també implica, en major o menor mesura, cert revisionisme sobre les afeccions que han de caure sota la categoria de la malaltia mental.

<sup>3</sup> Un treball excel·lent sobre la definició de la salut pot trobar-se en Lennart Nordenfelt (1993; 2007).

Començarem el nostre recorregut de la mà de la sociologia de la medicina, disciplina sorgida en els anys cinquanta del segle passat. Al llarg del nostre treball no solament ens servirem de les seues observacions en el camp de la psiquiatria, sinó que les reflexions que des de diferents àmbits s'han fet sobre la medicina i el concepte de malaltia corporal també ens seran útils per a il·luminar el nostre objecte.

#### LA MEDICINA COM A PRÀCTICA SOCIAL

Els primers treballs que posen significativament el focus en la dimensió social de la malaltia els podem trobar en la sociologia de la medicina que s'inaugura en els anys cinquanta amb el treball de Talcott Parsons (1951/2005). En el capítol que Parsons dedica a la pràctica moderna de la medicina, ja podem albirar una aproximació a la tríada de conceptes que hem identificat en la introducció. Partint d'una anàlisi sociològica funcional, en què la salut de l'individu és compresa com una necessitat per al funcionament correcte del sistema social, Parsons defineix la malaltia amb un peu en la biologia i un altre en la societat:

Podríem dir que la malaltia és un estat de pertorbació en el funcionament 'normal' de l'individu humà en la seua totalitat, incloent-hi l'estat de l'organisme com un sistema biològic, així com l'estat de les seues adaptacions personals i socials. Per tant, es defineix en part biològicament i en part socialment. La participació en el sistema social és sempre potencialment rellevant per a l'estat de la malaltia, per a la seua etiologia i per a les condicions d'una teràpia exitosa, així com per a altres coses (PARSONS 1951/2005, 290)

En aquesta primera aproximació a la definició de malaltia, veiem que Parsons no problematitza aquella *pertorbació en el funcionament normal* de l'organisme com a sistema biològic. No obstant això, per a comprendre el fenomen de la malaltia íntegrament, també cal atendre al seu caràcter social. L'anàlisi de Parsons se centra en aquesta part social de la malaltia, en el *rol de malalt* i en la configuració de la relació doctor-pacient. Estar malalt és entès com una forma de *desviació*, la característica diferencial de la qual resideix en la falta de responsabilitat: el malalt no és responsable per la seua condició, precisament per la dimensió orgànica en què aquesta s'ancora. El malalt s'ha desviat de les normes biològiques i també de les socials, però no se'n pot responsabilitzar perquè aquesta no ha sigut una elecció seua. Ara bé, encara que Parsons mantinga en el terreny natural la part orgànica del procés de la malaltia, ja adverteix que aquest àmbit no és impermeable, sinó que es veu també afectat per elements psicològics. Així doncs, mentre que hi ha una dimensió de la malaltia que es veu inafectada per

la intencionalitat humana, n'hi ha una altra de permeable a ella, com s'evidencia en el fet que les creences i els desitjos poden afectar processos fisiològics, com ara la recuperació d'un procés infecciós. Sense aprofundir més en la qüestió, Parsons reconeix que aquest problema es maximitza en el cas de la malaltia mental, en la qual els símptomes presentats no són orgànics, sinó comportamentals.

Podríem considerar la seua proposta com a híbrida, ja que en ella el fenomen de la malaltia no s'esgota en la dimensió biològica, ni tampoc en la social. Segons Simon Williams, la formació prèvia que Talcott Parsons tingué en el camp de la medicina va imbuir la seua anàlisi d'un cert to o qualitat realista, de manera que per a ell "la malaltia, en altres paraules, no és merament o simplement un estat de coses social, per molt important que aquests factors socials puguen ser" (WILLIAMS 2005, 126). Aquesta interpretació de la postura de Parsons no és compartida per tothom: el punt realista que Williams veu com un avantatge de la postura parsonsiana és, segons el parer de Stefan Timmermans i Steven Haas, poc menys que pesera o despreocupació intel·lectual: "la contribució de Parsons com a figura fundacional de la sociologia mèdica consistí a fitar l'àmbit social de la medicina, i deixà per a uns altres la biologia, fisiologia i patologia" (TIMMERMANS i HAAS 2008, 660). Aquesta tasca fou ben rebuda per les següents generacions de sociòlegs de la medicina, que abandonaran l'aproximació funcional per a apostar pel construccionisme social (BROWN 1995; EISENBERG 1988; FREIDSON 1970; TURNER 1987/1995), i donaran motiu a interessants debats en àmbits pròpiament filosòfics.

Abans de passar a problematitzar aquesta dimensió orgànica de la malaltia, ens endinsarem en el pensament d'alguns autors ben polèmics, com el conegut Thomas Szasz i —el tal vegada menys conegut— Hans J. Eysenck, els quals comparteixen aquesta visió realista, naturalista, de les malalties orgàniques i l'empren, precisament, per a qüestionar la realitat de la malaltia mental.

#### EL NATURALISME EN CONTRA DE LA PSIQUIATRIA

Usualment, es considera que les crítiques i revisions de la malaltia, física o mental, no provenen de les files naturalistes, sinó de les normativistes o construccionistes. Aquesta no és, però, una regla que es complisca sempre, n'és un bon exemple Thomas Szasz, l'autor del famós llibre *El mite de la malaltia mental* i personalitat destacada dins de l'anomenat Moviment Antipsiquiatria. Szasz denuncià l'ús de les institucions i pràctiques psiquiàtriques com a eina de control del comportament desviat i indesitjable: "Usar el terme 'malaltia mental' per a estigmatitzar, i així controlar, aquelles persones el comportament de les quals ofenga la societat (o el psiquiatra que en fa el 'diagnòstic') ha esdevingut costum" (SZASZ 1961/1972, 103).

Per a Szasz, la malaltia mental no és una malaltia real, sinó metafòrica. Segons ens narra, inicialment la classe *malaltia* contenia únicament elements que compartien la propietat comuna de fer referència a un estat fisicoquímic en què una estructura o una funció del cos humà es trobaven alterades. Posteriorment, elements addicionals foren inclosos en aquesta classe, però el criteri d'inclusió, atenent a Szasz, no fou llavors la presència d'alteracions orgàniques, sinó tan sols la presència de discapacitat i/o patiment. Així, fenòmens com la histèria o la depressió s'incloueren dins de la classe *malaltia*, encara que afegint-hi l'adjectiu *mental* per tal de diferenciar-los de les ocurrences corporals. Després, "els metges i especialment els psiquiatres començaren a anomenar 'malaltia' (és a dir, per descomptat, 'malaltia mental') a totes les coses en què pogueren detectar qualsevol signe de 'mal funcionament', sense importar en quina norma estiguera basat" (SZASZ 1961/1972, 58). Així, mentre que els primers elements de la classe s'hi incloïen per desviar-se de les normes naturals, els darrers s'hi incloüeren per desviar-se de les normes o preferències socials i individuals.

Amb un esperit similar, el psicòleg Hans J. Eysenck formula també una crítica a la psiquiatria perquè considera que, si bé aquesta té una part mèdica, majorment tracta problemes conductuals. Ni en el cas dels trastorns de personalitat, ni en les neurosis o psicosis s'hi troben lesions orgàniques, així com tampoc és possible correlacionar consistentment alteracions fisiològiques o bioquímiques. És per això que Eysenck considera que les classificacions psiquiàtriques no s'han elaborat basant-se en l'evidència científica, sinó que s'han realitzat atenent exclusivament a qüestions pràctiques. Igual que fa Szasz, aquest no posa en qüestió la legitimitat de la malaltia orgànica i considera que la psiquiatria no pot ser medicina perquè no tracta únicament malalties en el sentit orgànic:

Des del punt de vista del psicòleg, podríem tal vegada dividir la psiquiatria en la part mèdica, que tracta els efectes de tumors, lesions, infeccions i altres afeccions físiques, i la part conductual, que tracta trastorns del comportament adquirits a través dels processos ordinaris d'aprenentatge (EYSENCK 1960, 3)

Els trastorns conductuals es diferencien de les malalties, segons Eysenck, en almenys tres aspectes: en primer lloc, no s'ha identificat per a ells una única causa; en segon lloc, no és possible trobar una homogeneïtat entre ells que, a més, els diferencie clarament de les persones no diagnosticades; i, en tercer lloc, no existeix una cura específica per a ells. Per aquests motius, difícilment els podem aplicar el terme *malaltia*. En el seu lloc, hauríem d'abandonar aquesta concepció a favor de la noció de *dimensions de la personalitat*; i deixar en mans de la psicologia el seu tractament.

Allà on Eysenck va veure problemes dels processos normals d'aprenentatge, Szasz veu únicament *problems in living*, problemes de la vida: “expressions de la lluita de l'home amb el problema de com hauria de viure” (SZASZ 1970/1973, 21). No obstant això, per dir-ho fent servir la triada *disease, illness* i *sickness*, ambdós estarien d'acord que l'única malaltia que realment mereix anomenar-se així és la *disease*, la malaltia orgànica. Hem de comprendre bé Szasz quan diu que la malaltia mental és *un mite*: amb això no vol dir que no existisca en absolut, sinó que és un fenomen social en lloc d'un de natural; en cap cas hem d'entendre que les ocurrències socials i psicològiques a les quals va lligat no existeixen, sinó que han de ser enteses com a problemes pràctics i no com a malalties. Aquesta tesi no només contrasta radicalment amb la visió més estesa sobre la malaltia mental, sinó que posa en destret les seues implicacions pràctiques, ètiques i polítiques. Així, mentre que d'acord amb un punt de vista mèdic “la malaltia mental és com qualsevol altra malaltia i el tractament psiquiàtric, voluntari o no, és com qualsevol altre tractament” (SZASZ 1961/1972, 12), Szasz defensà que “la malaltia mental és un mite, la intervenció psiquiàtrica és un tipus d'acció social i la teràpia psiquiàtrica involuntària no és tractament, sinó tortura” (SZASZ 1961/1972, 12). Com ja avançarem en la introducció, la rellevància del debat entorn de la naturalesa de la malaltia mental té molt a veure amb les seues implicacions en el terreny pràctic, fonamentalment quan es considera que la malaltia pot ennuvoliar les capacitats cognitives i volitives de la persona que la pateix, de manera que es considere que aquesta perd la seua capacitat per a actuar de manera racional i lliure, la qual cosa pot portar-la a ser eximida de responsabilitat moral i legal, però també a un internament o tractament forçós, a més de l'estigma social i de l'impacte en la subjectivitat que suposa no considerar-se com un subjecte autònom: “de la manera com s'usen, generalment, els termes del diagnòstic psiquiàtric no descriuen entitats referents a una malaltia orgànica [*disease entities*], sinó que degraden i rebaixen la persona a la qual s'adscriuen” (SZASZ 1970/1973, 58).<sup>4</sup> En

<sup>4</sup> Ha de tindre's en compte que resulta impossible separar el posicionament de Szasz enfront de la psiquiatria de la seua postura radicalment liberal en el terreny polític. Per a Szasz, com a bon liberal, la llibertat individual representava un valor sagrat i qualsevol ingerència per part de l'àmbit col·lectiu o públic —això és, de l'Estat— hauria de ser contestada amb energia (SZASZ 1984); és per això també que aquest autor argumentarà en contra de la psiquiatria comunitària, alhora que aposta per una redefinició de la pràctica privada, deixant de pretendre que la psiquiatria és una branca de la medicina i acceptant obertament que tracta problemes de la vida. El subjecte que Szasz dibuixa és lliure i responsable de si mateix fins a extrems que poden resultar difícilment defensables; això és el que el porta a entendre la llibertat individual com una dura conquesta que ha de realitzar-se sense reserves enfront, tant de les adversitats externes com de les internes: “Perquè si la llibertat és la capacitat de prendre decisions sense reserves, llavors l'home naix encadenat. I el repte de la vida és l'alliberament. [...] En general, com més control guanya l'home sobre les seues condicions internes i externes, més lliure es torna; mentre que, si no aconsegueix aquest control, queda esclavitzat, o si, havent-lo guanyat, el perd, es converteix en esclau” (SZASZ 1970/1973, 1).

relació amb això, no solament l'obra de Szasz possibilita un pas important en el reconeixement dels drets dels pacients mentals, sinó que altres grans figures de l'anomenat Moviment Antipsiquiatria també van ser-hi elements clau; per exemple Ronald D. Laing (1960; 1967) —els treballs del qual sobre l'esquizofrènia com una resposta normal davant situacions adverses van ser acollits amb entusiasme— o Franco Basaglia (1967; 1968) —una figura clau en l'activisme contra l'internament dels pacients psiquiàtrics—. <sup>5</sup>

Seguint Bill Fulford i Werdie van Staden, per a Szasz —i el mateix pot ser extensible a Eysenck i altres figures de l'antipsiquiatria— “la legitimitat del concepte de trastorn mental depèn de la mesura en què el seu significat rastreja el del concepte de trastorn corporal” (FULFORD i VAN STADEN 2013, 389), ja que per a ell el concepte de trastorn o malaltia corporal no suposa problemes, en la mesura que apunta a una normalitat determinada pel món natural —per la integritat estructural i funcional— i no per les normes psicosocials, ètiques i legals. És per això que començàvem considerant Thomas Szasz com un pensador naturalista i revisionista: la malaltia mental no pot ser una malaltia real perquè no participa de la suposada objectivitat de la malaltia orgànica. A més a més, si en el futur descobrirem les causes orgàniques de la malaltia mental, aquesta deixaria de considerar-se *mental* i seria només *malaltia*. No obstant això, Szasz es mostra escèptic respecte a tal possibilitat: si bé es cert que en el passat s'han descobert algunes causes orgàniques per a fenòmens que eren considerats malalties mentals —com per exemple les psicosis associades a tumors cerebrals—, Szasz considera que l'esperança que tan sols siga *qüestió de temps* que trobem les bases o causes físicoquímiques o genètiques de la malaltia mental està essencialment equivocada, ja que en moltes ocasions l'única cosa que aquest concepte rastreja són comportaments sancionats socialment, raó per la qual podríem trobar-ne els correlats biològics, però mai trobar en la naturalesa la norma de la qual ens hem desviat.

Ara bé, no hem d'entendre que Szasz tan sols carrega les tintes contra la psiquiatria biològica, la que pretén concebre's com una branca de la medicina, sinó que igualment critica el model de la psicoanàlisi, en la mesura que aquest busca la mateixa legitimació a través de la il·legítima noció de malaltia mental. Atés que la psicoanàlisi és practicada com si la seua única meta fora traure el pacient d'un estat mental malalt per a retornar-lo a la salut mental, s'hi produeix

<sup>5</sup> Cal assenyalar, no obstant això, que hi ha qui considera que el factor que tingué major pes a l'hora d'aconseguir la desinternaltzació dels pacients psiquiàtrics no fou la pressió social i la força del Moviment Antipsiquiatria, sinó el fet de comptar amb fàrmacs relativament efectius per al tractament de les malalties mentals més greus, els neurolèptics (SHORTER 1997). El lector més avesat deu sospitar que, al seu torn, l'eficàcia dels neurolèptics també ha sigut qüestionada (WHITAKER 2009).

un encobriment de les qüestions de valor, el qual, segons el parer de Szasz, resulta paradoxal: “Mentre que és generalment acceptat que la malaltia mental té a veure amb les relacions socials o personals de l’home, es manté paradoxalment que els problemes sobre valors —és a dir, de l’ètica— no emergeixen en aquest procés” (SZASZ 1973, 20). A pesar que hui dia les postures crítiques amb la psiquiatria s’han suavitzat una mica —parlem de *psiquiatria crítica* més que d’*antipsiquiatria*— la veritat és que les denúncies sobre els encobriments dels valors en l’àmbit de la biomedicina i la biopsiquiatria continuen més que vigents, com veurem més endavant. (HERNÁEZ 2011; BERRIOS 2011).

Avancem ara cap a una altra gran figura en el debat al voltant de la malaltia mental. Ens quedem en el terreny del naturalisme, encara que aquesta vegada portarem a col·lació un defensor de la psiquiatria, que va fer front a les crítiques de Thomas Szasz, Hans Eysenck i altres figures de l’ansipsiquiatria: Robert Kendell.

#### EL NATURALISME AL RESCAT DE LA PSIQUIATRIA

En 1975, Robert Kendell va publicar “El concepte de malaltia i les seues implicacions per a la psiquiatria”, un article en què pretenia defensar la psiquiatria de les crítiques que havia rebut, les quals es concentraven al voltant d’un argument comú: que allò que els psiquiatres consideren malalties mentals no són malalties en absolut. El primer que Kendell subratlla és que per a posar en qüestió l’existència d’una malaltia o per a negar la seua realitat és necessari comptar ja amb un concepte de què és la malaltia; i el problema, al seu parer, és que tots aquests autors fan servir un concepte de malaltia que ja fou abandonat temps enrere no només per la psiquiatria, sinó per la medicina en el seu conjunt. Segons Kendell, tots els qui argumentaren que la psiquiatria no tractava malalties reals, ho feien basant-se, conscientment o inconscientment, en la premissa que algun tipus de lesió és essencial per a establir la presència de malaltia. Llavors, com que en el cas de les malalties mentals mai no s’hi ha demostrat cap lesió o alteració física, aquestes no són en realitat malalties.

Kendell ens narra que, des que en el segle XVIII es popularitzara la dissecció *post-mortem* dels cossos, la identificació de lesions físiques jugà un paper crucial en la descripció i determinació de les malalties; i més encara quan, a mitjan segle XIX, la invenció de potents microscopis va permetre observar patologies cel·lulars. Més recentment, els avenços tecnològics han permés expressar la malaltia en termes genètics, moleculars i de perturbacions biològiques estructurals. Així doncs, inicialment la malaltia va estar lligada



a l'anatomia patològica; i, més tard, també a les anomalies bioquímiques i fisiològiques. En aquest sentit, ha sigut comú pressuposar que la malaltia involucrava necessàriament alguna alteració física. A aquest enfocament se li suposa l'avantatge, com ja hem assenyalat, de capturar objectivament fets naturals, de manera que proporciona criteris objectius i fiables per a identificar la malaltia, la qual cosa és positiva per a previndre possibles abusos per part de la medicina. No obstant això, també presenta greus problemes.

D'acord amb Kendell, un dels problemes seria que aquells patiments la base física dels quals no s'haja identificat encara no poden ser considerats com a malalties, cosa que podria suposar un problema a l'hora de tractar-los en el si de la medicina, a pesar que poden suposar un important sofriment per a qui els pateix —i, per tant, semblen mereixedors d'atenció professional—. El segon problema és que no sempre que s'observa una alteració biològica estem enfront d'una malaltia: per exemple, aquest seria el cas de l'albinisme, en què a pesar que hi ha una manca d'un enzim involucrat en el metabolisme de la tirosina, no considerem que les persones albines patisquen per això una malaltia. El tercer problema que s'hi presenta, d'acord amb Kendell, és que, donada l'enorme variació que es dona fins i tot en individus sans en la mida, forma, composició química i eficiència funcional dels seus teixits i òrgans, no és possible establir clarament on acaba la variació normal individual i comença l'anomalia. Aquest problema el podem veure en el cas de la hipertensió o la diabetis: quin és el límit a partir del qual la pressió sanguínia o el nivell de sucre en sang no és normal?

Respecte a aquestes objeccions, voldríem realitzar algunes observacions. La primera d'elles fa referència al primer i segon problema identificat per Kendell en la concepció de la malaltia com a lesió; i és que s'hi fa patent, una vegada més, la vocació pràctica de la medicina: si no fora perquè hi ha condicions que considerem que són mereixedores d'atenció mèdica i unes altres que no impliquen cap perjudici o mal que haja de ser tractat mèdicament, no tindríem problemes de no classificar com a malalties aquelles condicions per a les quals no coneixem l'etiopatogènia, així com tampoc en tindríem a l'hora de classificar com a malalties afeccions com l'albinisme. El criteri operatiu per a incloure membres en la classe malaltia remet a factors socials. En aquest sentit, és interessant observar que les condicions que són considerades com a malalties sense conèixer-ne l'etiopatogènia poden formar part d'aquesta classe per les seues dimensions subjectives i socials (ERESHEFSKY 2009), és a dir, pel seu caràcter d'*illness* —sensació subjectiva de malaltia— i de *sickness* —pel reconeixement social del dret de qui les pateix a assumir el rol de malalt (PARSONS 1951/2005; WILLIAMS 2005)—. És per això que en aquests casos —els quals es coneixen en sociologia de la medicina com a *diagnòstics*

*controvertits* [*contested diagnoses*] (BARKER 2010; BROWN 1995)— els afectats “han de convèncer la institució de la biomedicina que el seu patiment és de caràcter mèdic” (BARKER 2010, 154).

Fem esment a això perquè, lluny de ser un fenomen marginal, suposa un repte per a la medicina contemporània, donat el fet que en les últimes dècades s’ha multiplicat el nombre de diagnòstics controvertits i de persones diagnosticades. Així, milions de persones “són diagnosticades amb una de les diverses síndromes caracteritzades per un conjunt de símptomes comuns, difusos i perturbadors, que poden anar des del dolor i la fatiga fins a trastorns del son i de l’ànim” (BARKER 2010, 153). Per esmentar-ne algunes, podem fer referència a les síndromes de la guerra del Golf, fibromiàlgia, fatiga crònica, intestí irritable i sensibilitat química múltiple. Algunes d’elles, com la síndrome de la guerra del Golf, són clarament competència de la psiquiatria, mentre que unes altres són disputades per considerar-se psicossomàtiques i sense unes competències i tractament clars.

Per últim, una observació a la tercera objecció plantejada per Kendell, el problema de la frontera entre la normalitat i la patologia. Segons Kendell, el fet que no puguem distingir clarament entre les variacions individuals normals i la patologia és una objecció contra el concepte de malaltia mental basat en la lesió orgànica. Les dificultats per a traçar els límits entre allò normal i allò patològic ja foren reconegudes antany en la medicina (CANGUILHEM 1943/2011; GIROUX 2010) i hui dia suposen un dels problemes més candents de la psiquiatria (EISENBERG 2007; FRANCES 2014; LANE 2007; HORWITZ i WAKEFIELD 2007). Ara bé, aquest problema no es presenta només en el cas de la malaltia com a lesió, sinó que igualment apareix en el cas de la malaltia com a disfunció biològica: és possible que allò normal i allò patològic es troben en un continu, en lloc de ser categories estanques. De fet, aquest enfocament *dimensional* — que desdibuixa les línies nítides entre allò normal i allò patològic i també entre els diferents trastorns, i els presenta en un continu o espectre— ha guanyat força en els últims temps (MELLOR MARSÀ i ARAGONA 2009).

Passem ara, després d’aquest *excursus*, a formular la proposta de Kendell, la qual pretén emmarcar-se dins del naturalisme basant-se en un model estadístic de la relació entre allò normal i allò patològic que trobe la seua base en la noció de *desavantatge biològic*: “Les malalties mentals es poden dir així justificadament perquè estan associades amb una reducció en la fertilitat i en l’esperança de vida, i això constitueix un desavantatge biològic” (KENDELL 1975, 313). Per a formular la seua proposta, Kendell acudeix al treball de John G. Scadding, que havia proposat que la clau per a definir el concepte de malaltia és si l’alteració subjacent situava l’individu en una posició de *desavantatge biològic*. Així, Scadding defineix la malaltia com:

la suma dels fenòmens anòmals mostrats per un grup d'organismes vius en associació amb una característica comuna específica o un conjunt de característiques per les quals diferisquen de la norma per a la seua espècie, de manera que això els situe en una posició de desavantatge biològic (SCADDING 1967, 877)

Ara bé, com Kendell assenyala, Scadding no elabora en què pot consistir aquell desavantatge biològic, i ací és on ell fa la seua proposta: en una reducció de la fertilitat i l'esperança de vida. Tenint la teoria de l'evolució com a marc, afirma que “en termes purament biològics la fertilitat és de summa importància, [perquè] determina quines espècies prosperaran i s'expandiran i quines s'extingiran, i quins genotips d'una espècie es convertiran en els dominants i quins continuaran sent estranys” (KENDELL 1975, 310-1). A més, la mort també té summa importància biològica, no solament per a l'individu, sinó també per a la seua descendència, particularment en animals socials com l'ésser humà, “la descendència del qual depèn dels seus progenitors durant una alta proporció de la seua vida” (KENDELL 1975, 311). Partint, així, d'aquesta definició de malaltia, Kendell es proposa, mitjançant l'evidència empírica, determinar si les malalties mentals són malalties reals o no. Per a això, acudeix a estudis sobre l'esperança de vida i la fertilitat de diferents grups considerats malalts mentals, i conclou que:

Hi ha evidència, per tant, que l'esquizofrènia i la depressió maníaca, juntament amb certs trastorns sexuals [homosexualitat] i diversos tipus de dependència a les drogues estan associats, bé amb una reducció en la fertilitat i/o en l'esperança de vida, i per aquesta raó són justificadament considerats com a malalties. El mateix podria demostrar-se cert finalment en alguns estats neuròtics i certs tipus de personalitat, però en el present l'evidència no és tan forta com per a justificar fermes conclusions en aquestes àrees (KENDELL 1975, 312)

Cal recordar al lector desprevengut que *homosexualitat* fou una categoria diagnòstica recollida fins a 1973 en el *Manual Diagnòstic i Estadístic dels Trastorns Mentals*, DSM, elaborat per l'American Psychiatric Association; i, d'una manera o altra, l'homosexualitat va romandre en el manual fins a l'any 1987, quan per fi fou totalment eliminada de les seues pàgines.<sup>6</sup> No ha de sorprendre'ns, doncs, que Kendell argumente a favor de l'homosexualitat

<sup>6</sup> En el DSM-II, el diagnòstic d'homosexualitat es va transformar en Pertorbació de l'Orientació Sexual —emfatitzant el fet que l'homosexualitat només era patològica si qui la patia desitjava ser heterosexual, la qual cosa va legitimar les teràpies de conversió sexual—. En la revisió següent, DSM-III, aquesta categoria va tornar a transformar-se, ara en Homosexualitat Ego-Distònica, i no seria fins a 1987, amb el DSM-III-R, que desapareixeria finalment del manual (DRESCHER 2015).

com a malaltia mental, encara que sí que podríem considerar que, el fet que el seu concepte de malaltia mental capture fenòmens com l'homosexualitat, suposa un argument en contra seua. En aquest sentit, Kendell no avança cap de les objeccions que els seus crítics podrien plantejar-li; en altres àrees, per contra, sí que especula sobre possibles crítiques que podrien ser-li formulades, com per exemple posar en qüestió la nítida distinció entre el que és un desavantatge purament biològic i un de social: podria argumentar-se, així, que els desavantatges existeixen, però que són conseqüència del fet que la gent siga etiquetada com a malalta i, per tant, no serien innats i inevitables. La rèplica de Kendell —al nostre judici, lluny de ser satisfactòria— consisteix a assenyalar que, en tal cas, l'entorn social hauria situat l'individu en una posició biològicament desavantajosa. A més, l'autor també fa front a una altra crítica anticipada: que hàgem de situar el focus sobre la supervivència de l'individu en lloc de fer-ho en la de l'espècie. En aquest sentit, podria argumentar-se per exemple que, en un entorn superpoblat, comptar amb individus homosexuals seria avantatjós per a l'espècie, fins i tot essent desavantatjós per a l'individu en qüestió. Si la rèplica anterior ens ha semblat insatisfactòria, la seua resposta final quasi fregaria la indiferència patològica:

La resposta, suggerisc, és que hem d'ignorar la importància creixent dels factors purament culturals a l'hora de determinar qui viu i qui mor; ignorar l'existència de fatals efectes de la discriminació en altres espècies, i també ignorar l'argument que és la supervivència del grup en lloc de la de l'individu el que importa (KENDELL 1975, 313)

La lleugeresa amb la qual Kendell despatxa aquestes crítiques és xocant. Tal vegada podríem hipotetitzar una explicació per a tal indiferència, i aquesta té a veure, novament, amb les implicacions pràctiques de la teràpia psiquiàtrica. Així com la crítica de Szasz era feroç perquè considerava que el tractament psiquiàtric era en moltes ocasions com una forma de tortura, Kendell considera, respecte dels esquizofrènics, que “si aquests foren simplement ignorats i no se'ls proveïra ni dels asils [*sanctuaries*] on poden ser alimentats i vestits, ni de la moderna farmacoteràpia, hi ha pocs dubtes que comparativament pocs arribarien a una edat avançada” (KENDELL 1975, 312). Igualment, allà on uns altres han vist institucions de control, Kendell considera que el propòsit primari dels asils del segle XIX fou el de protegir els dementats i només secundàriament la societat.

Per rellevant que siga aquesta discrepància entre Szasz i Kendell, no podríem concloure adequadament el repàs del pensament d'aquests dos autors sense subratllar que comparteixen una similitud gens menyspreable: el fet que

tots dos consideren que la legitimitat del concepte de malaltia mental depèn de la mesura en què el seu significat siga capaç de rastrejar el concepte d'alteració biològica. Així doncs, les discrepàncies a les quals arriben sobre el concepte de malaltia mental no es deuen sinó al fet que no comparteixen un mateix concepte de malaltia orgànica, contrastant així la visió d'alteració anatomicofuncional de Szasz amb la proposta de desavantatge biològic de Kendell (FULFORD i VAN STADEN 2013; KINGMA 2013). Mentre que Szasz va titllar la malaltia mental de mite perquè diferia de la malaltia física —en la mesura que únicament es desvia de normes socials i no biològiques—, Kendell va argumentar que, atès que la malaltia física i la mental eren *essencialment* el mateix, la malaltia mental era realment una malaltia.

Finalment, assenyalarem també que la postura de Kendell té implicacions revisionistes que el mateix autor reconeix i accepta, si bé molt més limitades que la de Szasz. L'aplicació del seu criteri de desavantatge biològic deixaria fora de la classe de malalties patiments com la neuràlgia postherpètica o la psoriasi; a més d'algunes afeccions que els psiquiatres tracten, com per exemple les neurosis. Reconèixer que alguns dels problemes amb els quals lúdia la psiquiatria no són malalties mentals no ha de portar-nos, segons el parer de Kendell, a no donar-los cobertura mèdica; per contra, de la mateixa manera que la medicina tracta fenòmens normals, com l'embaràs, la psiquiatria podria tractar fenòmens que no siguen patològics. És un deure dels psiquiatres, doncs, reflexionar sobre els límits de la seua competència i responsabilitat. En aquest sentit, la seua postura és similar a la de Szasz, que considera que la psiquiatria no hauria de desaparèixer com a especialitat —a excepció de la comunitària— sinó que hauria de reformular-se i acceptar honestament que tracta problemes de la vida que involucren qüestions axiològiques.

#### MÉS NATURALISME AL RESCAT

Robert Kendell no va ser l'únic que va reaccionar enfront dels atacs de l'antipsiquiatria. Una altra de les grans figures, coneguda també per les seues aportacions en la definició de la salut i la malaltia, és la de Christopher Boorse. En un article publicat en 1976, "Què hauria de ser una teoria de la salut mental", Boorse comença reconeixent els problemes que té la psiquiatria com a disciplina científica, ja que encara no gaudeix de consens en les seues bases teòriques. Segons el parer de Boorse, aquesta falta de consens no podrà superar-se fins que els psiquiatres i teòrics de la psiquiatria no abracen la fisiologia com el seu paradigma: el que no pot continuar fent la psiquiatria és emprar els conceptes de salut i malaltia de la medicina sense abraçar plenament

el paradigma mèdic; per tant, té dues opcions: o l'assumeix completament o, com van proposar els membres de l'antipsiquiatria, l'abandonen i redefeixen la disciplina.

Novament veiem la premissa que l'única malaltia real és la malaltia física. Per això, la psiquiatria només pot abraçar aquell model de malaltia; l'alternativa és abandonar completament el concepte. Fins ací, Kendell, Szasz, Eysenck i Boorse estan d'acord. Ara bé, Boorse, de manera semblant a Kendell, afirma la tesi —i amb això s'allunya de l'antipsiquiatria— que “la idea funcional de la salut en la medicina física s'aplica tan directament a la ment com al cos” (BOORSE 1976, 62). Boorse (1977, 555) dona les definicions següents per tal de formular la seua teoria bioestadística:

1. La *classe de referència* és una classe natural d'organismes de disseny funcional uniforme; específicament, un grup d'edat d'un sexe i d'una espècie.
2. Una *funció normal* d'una part o procés dins dels membres de la classe de referència és una contribució estadísticament típica d'aquesta a la seua supervivència i reproducció individual.
3. La *salut* d'un membre de la classe de referència és la *capacitat funcional normal*: la disposició de cada part interna a realitzar totes les seues funcions normals en ocasions típiques amb almenys l'eficiència típica.
4. Una *malaltia* és un tipus d'estat intern que deteriora la salut, és a dir, redueix una o més capacitats funcionals per davall de l'eficiència típica.

La clau de la seua teoria bioestadística és que les malalties són considerades com a interferències amb funcions naturals, i l'organització típica de les espècies, com un fet biològic; així, segons Boorse, el concepte de malaltia [*disease*] estaria lliure de valors. Aquesta concepció, a més, serviria no solament per al cas de la malaltia física, sinó igualment per a la malaltia mental, la qual “seria un cas especial obtingut centrant-se en les funcions de processos mentals” (1976, 63). Per tant, es podria parlar de salut i malaltia mental si hi ha cossos tals com funcions mentals; i, segons Boorse, encara que no es pot parlar de funcions mentals amb la mateixa especificitat que en el cas de les corporals —a causa de la plasticitat cerebral, el rol de la cultura i de l'entorn— sí que podem afirmar que certs tipus de processos mentals duen a terme funcions estàndard en el comportament humà, com per exemple la percepció o la memòria.

Boorse (1975), a més, distingeix entre dues dimensions de la malaltia, la malaltia com a *disease* i com a *illness*. Mentre que la primera és susceptible de ser definida sense involucrar qüestions morals, la segona sí que les inclou: en aquest sentit, la malaltia [*illness*] seria indesitjable per a qui la pateix, li conferiria dret a un tractament especial i, a més, seria una excusa per a un

comportament altrament criticable. Podem observar que, en aquesta concepció de malaltia com a *illness*, hi apareixen dues esferes que altres teòrics situarien en la dimensió social [*sickness*]: el dret a tractament i a ser eximit de responsabilitat moral i/o legal. Tal com assenyala Bill Fulford, “el seu objectiu [...] ha sigut demostrar que, si bé la pràctica mèdica pot estar carregada de valors, la teoria mèdica n'està lliure i, per tant, és científica” (FULFORD 2001, 80).<sup>7</sup>

Malgrat la intenció de Boorse de defensar la psiquiatria dels atacs patits, ell mateix també hi realitza algunes consideracions crítiques, encara que no pretenen ser una esmena a la totalitat: el retret que realitza té a veure amb el fet que els psiquiatres, en lloc d'investigar l'organització funcional normal de la ment humana per a conèixer en què consisteix la salut mental, arriben a formular els seus criteris sobre aquesta a través de tres rutes diferents (i inadequades): seleccionant trets de personalitat que són altament valorats —com la dignitat o la decència— i igualant-los amb la salut mental; pressuposant que certes condicions són malalties mentals —com les psicosis— i, per tant, els seus símptomes constitueixen una definició negativa de salut mental; o buscant criteris (socials) per a la definició d'un comportament normal com a mostra de salut mental. Al seu judici, “els mètodes utilitzats per la majoria dels teòrics de la ‘salut mental’ són essencialment indefensables” (BOORSE 1976, 81) i ací és on la psiquiatria es veurà forçada a decidir entre la investigació de l'organització funcional normal de la ment humana o l'abandó del camp de la salut en favor d'altres paradigmes.

Com era d'esperar, la proposta de Boorse també ha sigut objecte d'importants crítiques, les quals s'han dirigit, entre altres, a la seua noció de classes de referència i de funció biològica, així com a la seua afirmació que aquestes estan desproveïdes de qüestions de valor (FULFORD 2001; KINGMA 2007; 2013).

Boorse necessita justificar, sense referència prèvia a la salut i al trastorn, per què de totes les possibles formes d'agrupar els éssers humans, només l'edat, el sexe i tal vegada la raça són les agrupacions que sustenten una anàlisi de la salut i el trastorn. I com que Boorse es compromet a argumentar que la BST [teoria bioestadística] està lliure de valors [...] aquesta justificació per a admetre certes classes de referència també hauria de ser lliure de valors. Tal justificació no

<sup>7</sup> Un important aspecte filosòfic a problematitzar és la pretesa nítida distinció que se sol traçar entre fets i valors (PUTNAM 2002). La concepció ací presentada de la ciència com a encarregada de tractar únicament descripcions objectives, que ens remet a concepcions positivistes, presenta problemes importants, màximament quan del que es tracta és de despullar la medicina —disciplina clarament lligada a la pràctica— de tota qüestió de valor. A més, no solament és qüestionable la distinció mateixa, sinó que no és obvi que el fet que una classificació o empresa teòrica es trobe influenciada per valors impedisca que rastregi qüestions de fet; en aquest sentit, caldria preguntar-se per les condicions mateixes del coneixement i l'objectivitat (HASLANGER 2012).



circular i mancada de valors no és possible: restringir les classes de referència a grups humans sans seria circular, i altres propostes o bé no generen els resultats correctes o bé apel·len a valors (KINGMA 2013, 370)

Problemes similars es presentarien a l'hora de definir la funció mental, perquè sembla ser que l'arquitectura de la ment hi possibilita una aproximació pluralista, de manera que en última instància hàgem també de justificar per què hem triat una manera d'*esculpir els nostres conceptes* en lloc d'una altra. Ara bé, sembla que els valors ací implicats no són exactament de la mateixa naturalesa o no tenen el mateix rol que aquells valors que criticaven els membres del Moviment Antipsiquiatria. Del que ací es tracta és de triar el mapa conceptual que millor pugui representar l'arquitectura i el funcionament mental, sense perdre de vista que *millor* inclou també propòsits pràctics i valors, però no per això deixaria de capturar fets del món natural.

D'altra banda, filòsofs com Fulford han assenyalat que el concepte de funció suposa un altre repte per al naturalista, i és que no els pareix obvi que estiga exempt de valoracions, en aquest cas perquè no està clar que el seu ús sigui merament descriptiu: "perquè si bé és cert que *funció* és un terme propi de la biologia (que és una ciència), és un terme més aviat teleològic que (purament) causal; i la teleologia [...] pot connectar-se conceptualment a través de propòsits i intencions amb els valors" (FULFORD 2001, 83). Boorse defineix causalment *funció* com "una contribució causal estàndard a una finalitat efectivament perseguida per l'organisme" (BOORSE 1975, 57), en connexió amb les nocions evolutives de supervivència i reproducció. D'acord amb Fulford, la noció clau per a introduir valors en la teoria de Boorse és la de *finalitat* [*goal*]: així, el concepte de *finalitat*, com a concepte teleològic, ens permetria parlar d'un bon i mal funcionament depenent de si aquesta es pot aconseguir o no; però això només ocorre perquè el concepte mateix involucra un element avaluatiu: una finalitat és alguna cosa que es persegueix, es desitja, es necessita, etc.<sup>8</sup>

Per a superar algunes de les dificultats a les quals s'enfronta la teoria de Boorse, Jerome Wakefield proposà la seua teoria del trastorn com a *disfunció perjudicial*, en la qual la noció de funció es queda íntimament lligada a la selecció natural. Veurem aquesta proposta, considerada *hibrida*, al final d'aquest treball. Abans, però, farem un repàs a les postures que comprenen la malaltia, tant física com mental i en totes les seues dimensions, com un constructe social.

<sup>8</sup> El debat sobre la noció de funció biològica és molt extens en el camp de la filosofia de la biologia. Un recorregut per l'esmentat debat excedeix els propòsits d'aquest treball, però el lector interessat pot consultar *Psychiatry in the Scientific Image* (MURPHY 2006, 77-87), on Dominic Murphy, a més de rebutjar tant l'aproximació en termes evolutius de Boorse com la lectura en termes teleològics de Fulford, fa un repàs ràpid sobre les claus del debat al voltant d'aquesta noció.



Fins ara hem revisat postures que, malgrat que compten amb elements revisionistes, no han posat en qüestió la possibilitat de donar una definició i arribar a un coneixement lliure de valors de la malaltia orgànica [*disease*]. Referent a això, ja avançàvem en veure la postura de Talcott Parsons que aquesta crítica seria duta a terme des del camp del construccionisme social, mostrant com els factors socials són constitutius de la dimensió orgànica de la malaltia [*disease*], igual que es va fer amb les dimensions subjectives [*illness*] i socials [*sickness*].

Com Ian Hacking (1999) assenyala, usualment l'objectiu social de les anàlisis construccionistes socials sol ser el d'emfatitzar els aspectes culturals i històrics de classificacions i d'objectes comunament compresos com a naturals; d'aquesta manera, se'ls despulla de la necessitat associada al món natural i se n'exposa la contingència, la qual cosa en possibilita la revisió i fins i tot la possible eliminació. Seguint també Hacking, introduïrem superficialment una distinció que ell traça per a comprendre millor els discursos sobre la construcció social: la construcció d'idees i d'objectes. Així, la construcció d'idees remet al fet que les nostres classificacions, el nostre coneixement, estan travessats per factors socials; en canvi, en parlar de la construcció de l'objecte s'apunta al fet que la seua pròpia naturalesa és social i no natural. Una vegada feta aquesta distinció, podem veure com alguns teòrics de la construcció social argumentaran a favor de la construcció de la idea de malaltia [*disease*], però no de l'objecte. Aquestes seran les formes de construccionisme social de la malaltia més febles. En canvi, els autors que aposten per formes construccionistes socials més radicals, o bé senzillament no traçen la distinció entre idea i objecte, o bé afirmaran que tant la idea de malaltia com la malaltia com a objecte estan construïts socialment. El sociòleg Eliot Freidson, que pertany al primer grup, argumenta a favor de la construcció social de la idea de malaltia, però es mostra més caut a l'hora de parlar de la seua construcció com a objecte, deixant espai a la possibilitat que, després de tot —a pesar que les nostres classificacions estan travessades per factors socials— aquest concepte aconseguisca capturar alguna qüestió de fet, un estat natural de coses:

la malaltia com a tal pot ser una malaltia biològica [*biological disease*], però la idea de malaltia no ho és, i tampoc ho és la forma en què els humans responen a ella. Així, la desviació o malaltia biològica es defineix socialment i està envoltada d'actes socials que la condicionen (FREIDSON 1970, 209)

Quan Freidson parla de la malaltia *com a tal*, podem interpretar fàcilment que s'està referint a la malaltia com a objecte contràriament a la malaltia com

a idea, a la qual fa al·lusió seguidament. Traçant aquesta distinció, Freidson no qüestiona l'existència d'una dimensió de la malaltia que siga independent dels factors socials —independentment real—, però quan l'ésser humà interactua amb la malaltia, no només ho fa amb l'objecte, sinó també amb la idea:<sup>9</sup> “mentre que la malaltia com a estat biofísic existeix independentment del coneixement i l'avaluació humana, la malaltia com a estat social és creada i modelada pel coneixement i l'avaluació humana” (FREIDSON 1970, 223). Ara bé, encara que tals estats biofísics siguen independents del nostre coneixement sobre ells, el fet que nosaltres els considerem malalties —això és, considerem que s'han desviat d'una norma— no és merament un fet natural, sinó que involucra qüestions de valor perquè els considerem indesitjables.

Christopher Boorse es va referir a aquest tipus de propostes com *normativistes febles* —en un sentit similar l'esmentat més amunt— ja que tenen un component descriptiu i un altre de normatiu, els quals deixen espai per a una dimensió natural, una dimensió biofísica, fora del discurs i de les pràctiques socials; tot això en contrast amb el normativisme fort, el qual opina “que els judicis de salut són avaluacions pures sense significat descriptiu; el normativisme feble permet que tals judicis tinguin un component tant descriptiu com normatiu” (BOORSE 1975, 51). Dins d'aquest tipus de normativisme feble, trobaríem també autors com H. Tristram Engelhardt (1976) —que realitzà una fina anàlisi sobre els diferents tipus de valors que entren en joc en les diferents dimensions de la malaltia, així com en la malaltia com a estat biofísic i les explicacions que realitzem sobre aquest estat—; o Joseph Margolis, que mantingué un interessant debat amb Christopher Boorse sobre la imbricació dels valors en la determinació de la funció natural (BOORSE 1975; MARGOLIS 1959).

Passem ara a explorar les postures que aposten per una construcció forta, pròpies d'autors que, o bé no tracen la distinció entre la construcció d'idees i d'objectes, o bé consideren que el vessant orgànic de la malaltia està construït com a idea i com a objecte. Aquest seria el cas d'autors com Peter Sedgwick (1982) o Bryan Turner (1984/2008; 1987/1995), que al seu torn recorren amb freqüència als estudis del filòsof francès Michel Foucault (1963; 1966). Així, mentre que una postura construccionista feble en el terreny de la malaltia

<sup>9</sup> Ací, Freidson sembla estar apuntant al que, anys més tard, Ian Hacking (1996) va encunyar com *efecte bucle* i Sally Haslanger (2012) com *construcció social discursiva*; l'impacte que el discurs té, tant en la subjectivitat i comportament humans, com en la creació o configuració d'objectes: “quan un metge diagnostica la condició d'un ésser humà com a malaltia, en canvia el comportament mitjançant el diagnòstic: un estat social s'afig a un estat biofísic assignant el significat de malaltia a la malaltia. És en aquest sentit que el metge crea malaltia de la mateixa manera que el legislador crea crim, i que la malaltia és una espècie de desviació social analítica i empíricament diferent de la mera malaltia” (FREIDSON 1970, 223).

orgànica desemascararia la construcció social del coneixement mèdic —en termes de Hacking, la construcció de la idea o classificació de la malaltia—, la visió forta del construccionisme social de la malaltia sostindria no solament que el nostre coneixement mèdic es troba construït, sinó que la malaltia, com a entitat, també es troba construïda: que està constituïda per factors socials en lloc d'estar-ho per fets naturals. Així, encara que totes dues postures estarien d'acord que “el coneixement mèdic sobre la malaltia [*disease*] no és necessàriament de naturalesa objectiva, sinó que és construït i desenvolupat pels seus autors i les parts interessades, que amb freqüència tenen una agenda avaluativa sòlida” (CONRAD i BARKER 2010, 76), els segons farien un pas en el terreny ontològic. Vegem com ho presenta un defensor de la construcció social radical, Bryan Turner.

Turner està particularment interessat en la relació entre el coneixement i el poder en la distribució social de la salut i la malaltia. El seu màxim interès resideix en l'estudi del rol del discurs mèdic com a fonament del seu poder. La seua aproximació involucra una anàlisi de la idea de malaltia com a interpretació mèdica de la realitat, la qual cosa el portarà a realitzar una avaluació crítica de la disposició d'allò natural a través de la perspectiva de la sociologia del coneixement de les categories de *disease* i *illness*. Així, des d'una perspectiva sociològica i filosòfica, tractarà els conceptes de la ciència mèdica com a producte de canvis culturals, sempre constrets pel llenguatge. Turner es basa fortament en el treball de Foucault, que veia el coneixement científic com qualsevol altra forma de coneixement humà: constret per les limitacions del llenguatge, constituït per interpretacions i metàfores i lluny d'una pretesa objectivitat. Segons Turner, això té conseqüències importants per a la medicina:

No podem continuar considerant les malalties com a esdeveniments naturals del món que ocorren fora del llenguatge amb què se'ls descriu. L'entitat de la malaltia és producte dels discursos mèdics, que al seu torn reflecteixen la manera dominant de pensar d'una societat. [...] El que les coses són depèn de com són definides; com es defineixen les coses depèn de com la cultura general situa els fenòmens dins dels espais de convenció. Si adoptem aquesta teoria del coneixement, llavors la malaltia [*disease*] no és una entitat patològica en la naturalesa, sinó el resultat de processos sociohistòrics (TURNER 1987/1995, 11-2)

Turner afirma estar fent *ontologia social* en preguntar-se què és la malaltia [*disease*]; i tenia per labor dilucidar les *constitucions socials* de les malalties [*diseases*].<sup>10</sup> A més, l'anàlisi de l'ontologia social de la malaltia ens ha de

<sup>10</sup> Estem emfatitzant entre claudàtors l'ús del terme *disease* per a no perdre de vista que el que està en qüestió és la dimensió orgànica, com a fenomen biològic o natural, de la malaltia, tant física com mental.

portar a considerar també l'ontologia del cos, perquè un concepte no es pot comprendre sense l'altre; igual que tampoc pot comprendre's la separació de la medicina corporal i la medicina mental —psiquiatria— sense entendre la vella escissió que ha dominat nostra cultura, l'escissió entre cos i ment. L'estudi sociològic i filosòficoontològic del cos és fonamental, al parer de Turner, per a entendre el fenomen de la malaltia, atès que la malaltia [*disease*] es presenta com una entitat neutral i natural que resideix en la naturalesa, és a dir, en el cos del pacient. Turner, així com altres construccionistes, formularà la pregunta ontològica per a la realitat del cos i també respondrà a la pregunta a través del construccionisme social, apuntant als processos i realitats socials que el conformen (BUTLER 2011; TURNER 2008; WILLIAMS 2006).

La rellevància que tenen aquests discursos per al nostre treball és que, com hem vist fins ara, bona part dels intents de donar compte de la malaltia mental —siga per a legitimar-la o per a deslegitimar-la— tenen a veure amb la consideració d'aquesta com a malaltia corporal, atès que la noció de malaltia corporal sembla habitualment menys problemàtica, en considerar-se que únicament està rastrejant un fet natural quan traça la línia entre allò normal i allò patològic, de manera objectiva, sense que hi medien qüestions de valor. Això seria, segons el parer de Peter Sedgwick, un greu error:

Molt encertadament, els antipsiquiatres han assenyalat que les categories psicopatològiques es refereixen a judicis de valor i que la malaltia mental n'és una desviació. D'altra banda, les crítiques antipsiquiàtriques s'equivoquen quan s'imaginen que la medicina física és essencialment diferent, en la seua lògica, de la psiquiatria. Un diagnòstic de diabetis, o parèsia, inclou el reconeixement de normes o valors (SEDGWICK 1982, 38)

Mostrar la construcció social de la malaltia orgànica és exposar que aquesta no es desvia de normes biològiques, sinó socials, i que, a més a més, sols està constituïda per factors socials; així mateix, que necessàriament està travessada per les contingències i arbitrarietats de cada cultura —i, per aquesta raó, la malaltia no pot ser un fenomen universal ni necessari, sinó que albergaria espai en el seu si per a la revisió i reconstrucció—. Arribats a aquest punt, ens agradaria detindre'ns per a analitzar amb una mica més de detall algunes tesis i arguments construccionistes, a fi de precisar —seguint l'esperit de Hacking— quins són els diferents sentits en què algú podria considerar un objecte com a constructe social; per a fer-ho, recorrerem a algunes distincions traçades per la filòsofa Sally Haslanger (2012).

ALGUNES OBSERVACIONS SOBRE LA CONSTRUCCIÓ SOCIAL  
DE LA MALALTIA (MENTAL)

El primer que podem començar clarificant, encara que ja ha sigut expressat anteriorment, és el sentit general de construcció social: “quelcom és una construcció social en el sentit genèric si, i només si, és un producte intencionat o inintencionat d’una pràctica social” (HASLANGER 2012, 86). Per a precisar de quina manera un objecte pot construir-se socialment, Haslanger (2012, 87-8) ens ofereix les següents distincions:

Construcció causal: alguna cosa està causalment construïda si, i només si, els factors socials hi juguen un rol causal en portar-la a l’existència o, en una mesura substancial, siga com és.

Construcció constitutiva: alguna cosa està constitutivament construïda si, i només si, en definir-la hem de fer referència a factors socials.

Construcció discursiva: alguna cosa està discursivament construïda només en cas que siga com és, en una mesura substancial, pel que se li atribueix (i/o autoatribueix).

A través d’aquestes eines analítiques podem ampliar la nostra comprensió d’algunes qüestions tractades més amunt; màximament quan s’entrellacen amb les dimensions orgànica, subjectiva i social de la malaltia. Ara repassarem, amb la seua ajuda, el debat exposat fins ací. En primer lloc, hem començat en els anys cinquanta amb la proposta del sociòleg Talcott Parsons, que ja feia servir la distinció entre les dimensions orgànica, subjectiva i social de la malaltia. Segons l’anàlisi de Parsons, les dimensions subjectiva i social de la malaltia estan socialment construïdes constitutivament; així doncs, per a definir-les, és necessari apel·lar a factors socials: tant per a comprendre la subjectivitat del malalt com per a comprendre’n el rol cal atendre a valors i pràctiques socials, i no merament a fets naturals. La dimensió orgànica de la malaltia, per part seua, no estaria constitutivament construïda socialment, perquè en la seua definició no seria necessari fer referència a factors socials; sí que estaria, en canvi, construïda causalment, perquè els factors socials —com per exemple les estructures i pràctiques mèdiques— tindrien una incidència en el seu curs; fins i tot la mateixa subjectivitat, la disposició amb la qual s’enfronta la malaltia, hi pot incidir, tal com s’ha mostrat en el cas de la recuperació enfront de processos infecciosos i, encara més, en el cas de les malalties mentals. En el cas de les dimensions subjectiva i social de la malaltia, hi observem també el rol de la construcció discursiva: tant la subjectivitat del malalt com el rol social que aquest adopta són configurats, en bona mesura, pel discurs mèdic.

Per part seua, l'anàlisi de naturalistes com Thomas Szasz apuntaria que les malalties corporals no es troben constitutivament construïdes socialment —en la mesura que apunten a fets naturals, a alteracions fisiopatològiques—; en canvi, això sí que s'esdevé en les malalties mentals, perquè tot el que trobem en la seua definició són factors socials, i no fets naturals. La seua construcció també seria discursiva, perquè és el discurs mèdic —al costat de l'avaluació social negativa— el que les fa existir. De la mateixa manera, aquest discurs mèdic impacta i conforma tant la subjectivitat individual com el rol social del malalt. Altres postures naturalistes, com la de Robert Kendell, considerarien que ni la malaltia mental, ni la malaltia corporal estan constitutivament construïdes, perquè es poden trobar fets naturals, en forma de disfunció biològica, que les defineixen, sense que hi medien qüestions de valor o factors socials de cap mena. Per part seua, Christopher Boorse, en distingir clarament entre el vessant orgànic de la malaltia [*disease*] i el vessant subjectiu i social (recollides totes dues sota el concepte d'*illness*) deixa espai a la construcció social constitutiva i causal de les dimensions subjectiva i social —perquè es veuen impactades pel discurs mèdic i recullen també les valoracions, preferències i el significat que els atorga tant qui les pateix com la resta de la societat—; però deixa un espai lliure de construcció per al vessant orgànic de la malaltia [*disease*], per a la definició de la qual no cal atendre a factors socials, sinó merament a la funció biològica, a través de la seua anàlisi bioestadística. Finalment, les formes més radicals de construccionisme social assenyalarien la construcció social causal i constitutiva de les dimensions orgànica, subjectiva i social de tota malaltia, física o mental. En el cas que acabem de revisar en la secció anterior d'aquest treball, la construcció social de la malaltia corporal, hem vist que l'argumentació de Bryan Turner s'ha centrat a revelar com la malaltia corporal està constitutivament construïda de manera radical —ja que, per tal de definirla, és imprescindible apel·lar a factors socials, en aquest cas, a relacions de poder; fins al punt que amb això esgotarem la seua definició—. Quan Bryan Turner afirma que la malaltia [*disease*] és producte dels discursos mèdics no està afirmant altra cosa que, allò que semblava remetre a una entitat de la naturalesa, no remet sinó a alguna cosa que no siga el mateix discurs, sustentat pel poder mèdic. Per tant, els factors socials —en aquest cas, el discurs i les pràctiques mèdiques— constitueixen —en un sentit ontològicament fort— la malaltia [*disease*].

Avaluem, per últim, l'argument que Turner ens ofereix, abans de passar a revisar les postures híbrides. El seu argument es pot condensar en les línies següents: (1) totes les narratives —conjunts d'esdeveniments dins d'un llenguatge— depenen de diverses convencions del llenguatge —que és un sistema autoreferencial—; (2) res ocorre fora del llenguatge; (3) per tant, el

que sabem “del món” és simplement el resultat de convencions arbitràries que adoptem per a descriure el món; (4) si adoptem aquesta teoria del coneixement, llavors la malaltia [*disease*] no és una entitat patològica en la naturalesa, sinó el resultat de processos sociohistòrics (TURNER 1987/1995, 11-2).

Encara que aquest argument sembla funcionar en un pla epistèmic, en observar-lo detingudament podem apreciar que no és exactament així: en primer lloc, alguns elements són ambigus, com per exemple la segona premissa, que sentència que res *ocorre* fora del llenguatge —la qual cosa sembla remetre més a un pla metafísic—; en segon lloc, i més important, el quart punt ens remet directament a un pla ontològic. Realitza un pas —il·legítim, al nostre judici— des de la teoria del coneixement a l'ontologia. Com hem assenyalat al principi, la intenció de Turner era la de realitzar una ontologia social; i ara veiem clarament com ho du a terme, perquè el seu argument sobre la realitat de la *disease* col·lapsa els plans epistèmic i ontològic inferint el que la malaltia *és* a través de com la coneixem: “El que les coses són depèn de com són definides” (TURNER 1987/1995, 11). Sense entrar a valorar l'encert o no de la teoria del coneixement proposada —i basada en Foucault—, podem criticar aquest pas il·legítim d'una esfera a una altra. L'argument proposat ací segueix la tendència d'aquells que desitgen presentar la realitat no com a objecte dels nostres esforços epistèmics, sinó com el seu producte, passant del reconeixement dels factors socials que condicionen el nostre coneixement a la tesi que tals factors en constitueixen la realitat. Acceptar que la realitat només pot ser representada pel llenguatge no ens porta a concloure que la realitat està constituïda pel llenguatge o que fora d'aquest *no ocorre res* (SEARLE 1995); de la mateixa manera, acceptar que els nostres criteris epistèmics poden estar influenciats per factors i relacions socials no iguala el contingut de les nostres atribucions amb aquells (HASLANGER 2012).

#### DE L'ANTIPSIQUIATRIA A LA PSIQUIATRIA CRÍTICA

L'últim tram del nostre recorregut pels debats entorn de la definició de la malaltia mental arriba fins als nostres dies, de la mà del que s'ha denominat *psiquiatria crítica*. Igual que en les dècades passades, quan els debats tenien un rerefons social molt important, hui dia els problemes teòrics de la psiquiatria també influencien i es veuen influenciats per l'opinió pública i els problemes socials. Mentre que en els seixanta i setanta el més preocupant era la qüestió de l'internament dels pacients mentals —és a dir, el més preocupant eren les malalties mentals greus—, hui dia la major preocupació apareix en el si dels trastorns menys severs —és a dir, en el que s'ha denominat *expansió del diagnòstic*

*psiquiàtric i processos de patologització de la vida quotidiana* (BALLESTEROS 2018A; FRANCES 2014). Actualment, és comú trobar postures que consideren que la psiquiatria té legítimament un nucli d'afeccions mentals per tractar, però aquestes no representarien el gruix de la pràctica psiquiàtrica. A més, bona part del debat té també relació amb el tractament psicofarmacològic de la malaltia o trastorn mental, el qual s'ha imposat en la pràctica psiquiàtrica i ha ajudat a forjar una imatge biologicista de la ment i les patologies mentals (BALLESTEROS 2018B; KRAMER 1993). En aquest sentit, hi apareixen també veus crítiques que qüestionen l'eficàcia del tractament farmacològic i posen en dubte la imatge de progrés que amb freqüència es presenta de la psiquiatria (BENTALL 2009; KIRSCH 2010; WHITAKER I COSGROVE 2015).

Aquests canvis en el focus de les problemàtiques té molt a veure amb el desenvolupament mateix de la psiquiatria. Encara que no tenim espai aquí per a desenvolupar aquesta qüestió en profunditat, ens agradaria apuntar una sèrie de condicions que han afavorit aquest procés expansiu de la psiquiatria. En primer lloc, l'aparició del DSM-III en 1980 se situa com una fita atès que, en oferir definicions operacionalitzades per als diferents trastorns mentals, aconseguí un consens mai vist abans en el diagnòstic psiquiàtric (FRANCES 2014), la qual cosa va ser molt positiva per a respondre a les veus més crítiques: era, almenys, el primer pas per a portar la psiquiatria al terreny segur de la medicina, ja que se'n va aconseguir estandarditzar els criteris diagnòstics, independentment de l'escola a la qual pertanguera el terapeuta. El problema del DSM-III —que s'ha vist agreujat posteriorment— és que ja contenia criteris diagnòstics laxos, mentre que a penes reconeixia la importància d'atendre al context per a avaluar si unes símptomes podien ser-hi una resposta normal o no. En segon lloc, l'aparició del *Prozac* a la fi de la dècada dels 80 suposà també una revolució, perquè per fi es comptava amb una medicació específicament dissenyada per al tractament d'un trastorn mental —la depressió—, la qual va poder ser consumida per milions de persones a causa dels seus (relativament) pocs efectes secundaris (KRAMER 1993). En tercer lloc, cal destacar explícitament el rol de la indústria farmacèutica a l'hora de pressionar per al consum dels seus psicofàrmacs (KIRSCH 2010; LANE 2007). En quart lloc, cal recordar que la dècada dels 90 és usualment coneguda com “la dècada del cervell”, a causa de l'enlairament de les neurociències: en aquest sentit, donada la relació de legitimació que hem exposat al llarg de l'article entre la biologia i la malaltia mental, fou d'una importància crucial començar a trobar els primers correlats entre les símptomes mentals i les seues bases biològiques (ANDREASEN 1995). Finalment, tampoc no podem obviar la pròpia estructura dels nostres serveis sanitaris, en què preval l'atenció psiquiàtrica en lloc de la psicològica.



Ho posem sota el punt de vista d'Allen Frances, que va estar al capdavant de la redacció del DSM-IV i ara entona el *mea culpa*:

La promoció de malalties no pot ocórrer en un buit: exigeix que les companyies farmacèutiques busquen la col·laboració activa dels metges que escriuen les receptes, els pacients que les demanen, els investigadors que inventen els nous trastorns mentals, els grups de consumidors que advoquen per més tractament i els mitjans de comunicació i Internet que difonen el missatge (FRANCES 2014, 29)

En aquest paradigma d'expansió psiquiàtrica, una de les qüestions més apressants és la de distingir els trastorns *reals* del que són *afeccions normals patològitzades*; màximament quan els criteris diagnòstics s'han vist usualment rebaixats edició rere edició del DSM, de manera que engloben cada vegada més persones (FRANCES 2014). Ara bé, com podem distingir quines afeccions són realment trastorns mentals i quines no ho són, quan encara no hem arribat a un consens sobre què és o ha de ser un trastorn mental? Davant la dificultat per a respondre a això, psiquiatres com Frances aposten directament per una solució pragmàtica: traçar la línia entre allò normal i allò patològic de manera que englobe només aquells que presenten símptomes greus i que no remetran per ells mateixos amb el pas del temps:

El trastorn psiquiàtric consisteix en símptomes i conductes que no s'autocorregeixen, és a dir, en una interrupció del procés homeostàtic normal de curació. La inflació diagnòstica ocorre quan confonem les típiques pertorbacions que són part de la vida de tots amb vertaders trastorns psiquiàtrics (FRANCES 2014, 32)

Aquest problema d'inflació diagnòstica el trobem, entre altres, en afeccions com el Trastorn Depressiu Major (BALLESTEROS 2018A; HORWITZ I WAKEFIELD 2007; WAKEFIELD 2015), el Trastorn d'Ansietat Social (LANE 2007) o el Trastorn per Dèficit d'Atenció i Hiperactivitat (EISENBERG 2007). La gravetat de la situació resideix, no solament en la inflació diagnòstica, sinó en el tractament psicofarmacològic d'aquestes condicions quan no són *realment* patològiques, ja que, a part de comportar efectes secundaris, poden interferir amb els processos homeostàtics de curació.

Precisament, a dur a terme aquesta tasca de distingir allò realment patològic d'allò *patològitzat* s'ha centrat Jerome Wakefield, amb una profunda anàlisi sobre el fenomen de la depressió i la progressiva transformació de la tristesa intensa en depressió que hem viscut recentment en la nostra societat. Segons el parer de Wakefield, el fenomen que recullen els manuals diagnòstics

no és un trastorn mental, sinó que els criteris diagnòstics són tan laxos i mancats de referències al context que confonen sistemàticament el que és una resposta normal, adaptativa, dels mecanismes de superació de la pèrdua, amb un funcionament patològic d'aquests (BALLESTEROS 2018A; HORWITZ i WAKEFIELD 2007). Exposarem a continuació el seu concepte de trastorn mental per a comprendre millor la seua crítica.

#### EL TRASTORN MENTAL COM A DISFUNCIÓ PERJUDICIAL

Abans de presentar la proposta de Wakefield, convé ressaltar un canvi important en la seua terminologia; i és que Wakefield abandona el terme *malaltia* en totes les seues dimensions (*disease*, *illness* i *sickness*) per a substituir-lo pel de trastorn [*disorder*]. Segons el seu parer, el terme *trastorn* és més ampli i cobreix tant les lesions traumàtiques com les dimensions orgàniques [*disease*] i subjectives [*illness*] de la malaltia. No obstant això, Wakefield comparteix amb Kendell i Boorse la seua aproximació naturalista a través de la noció de funció biològica, encara que amb alguns refinaments. Allí on Boorse va haver de servir-se de la noció de classe de referència per a poder fonamentar objectivament el concepte de malaltia, Wakefield hi assenyala el *disseny*<sup>11</sup> evolutiu per formular les nocions de funció i disfunció sense introduir qüestions de valor.

Mentre que per a Boorse el concepte de malaltia [*disease*] estava lliure de valors perquè era igual al concepte de disfunció, per a Wakefield el concepte de trastorn [*disorder*] inclou ja valors, en tant que és la suma d'una disfunció i una valoració negativa d'aquesta; per aquest motiu, batejà la seua teoria sobre el concepte de trastorn com *anàlisi de la disfunció perjudicial*. Vegem-ho més detalladament:

un trastorn es troba en el límit entre el món natural donat i el món social construït; un trastorn existeix quan el fracàs dels mecanismes interns d'una persona en l'acompliment de les seues funcions segons disseny natural afecta perjudicialment el benestar de la persona segons allò definit pels valors i significats socials (WAKEFIELD 1992, 373)

Wakefield defineix la seua proposta com a *híbrida* en tant que el terme *disfunció* és un terme científic, propi de la biologia evolutiva, mentre que el terme *perjudicial* apunta a estàndards socioculturals; d'aquesta manera, el

<sup>11</sup> Òbviament, quan parlem de disseny o quan més endavant parlem de propòsit no estem fent sinó un ús metafòric dels termes, ja que aquest tipus d'intencionalitat no pot ser atribuïda als processos naturals.

concepte de trastorn mental tindria una naturalesa híbrida. Quan parla de mecanismes interns, aquests han d'entendre's de manera àmplia, fent referència tant a estructures físiques i òrgans com a estructures i disposicions mentals. En aquest sentit, pot parlar de mecanismes motivacionals, cognitius, afectius o perceptius. Igual que Boorse, Wakefield considera que els processos mentals exerceixen un rol en la supervivència i la reproducció, de manera que —com Darwin també emfatitzava— aquests foren seleccionats naturalment per la funció que exercien. Així, la disfunció apareix quan “algunes condicions mentals interfereixen amb la capacitat d'aquests mecanismes mentals per a realitzar les funcions per a les quals van ser dissenyats” (WAKEFIELD 1992, 375). És important notar que una aproximació evolutiva és plenament compatible amb una descripció en termes mentals o comportamentals dels mecanismes com a descripcions biològiques, la qual cosa és clau per a no caure en un paradigma tan limitat com el que plantejava Thomas Szasz.

Per a definir amb més detall la seua noció de funció biològica, Wakefield hi estableix un paral·lelisme amb la funció dels artefactes artificials: així com la funció d'un artefacte és important perquè a través de la seua connexió amb el disseny i el propòsit posseeix un gran valor explicatiu, els mecanismes naturals poden ser —parcialment— explicats inferint-ne els efectes, de manera que les funcions naturals són aquelles els efectes de les quals apareixen en tals explicacions. Així, “una funció natural d'un òrgan o un altre mecanisme és un efecte de l'òrgan o mecanisme que entra en una explicació de l'existència, estructura o activitat de l'òrgan o mecanisme” (WAKEFIELD 1992, 386). La teoria que hui dia millor pot oferir-nos aquest tipus d'explicacions sobre la presència i estructura de tals mecanismes és la teoria de l'evolució: aquells mecanismes que tingueren un efecte positiu sobre l'èxit reproductiu d'organismes passats durant suficients generacions hi van ser seleccionats naturalment, de manera que perduren en els organismes actuals. Havent construït d'una manera científica, objectiva, la noció de funció, la seua contrària —la disfunció— es presenta també com un concepte científic. Una dificultat que apareix, com així ho reconeix Wakefield, és que arribar a tindre coneixement sobre quines són, de fet, les funcions dels mecanismes mentals és una tasca de gran complexitat i, previsiblement, subjecta a grans controvèrsies; sobretot considerant la gran ignorància que hi tenim actualment.

Wakefield dedica la major part dels seus esforços argumentatius a aquella primera part de la definició de trastorn, a la disfunció; no obstant això, també fa algunes observacions sobre la segona condició: que una disfunció ha de ser perjudicial per a ser autènticament un trastorn. En aquest sentit assenyala que, atès que l'entorn actual difereix tant de l'entorn en què operaven les pressions selectives, és possible que les afeccions que anteriorment podrien haver suposat

un desavantatge o perjudici, hui dia no suposen un problema. És per això que no n'hi ha prou amb l'existència d'una disfunció per parlar de trastorn: aquest, a més, ha de suposar un perjudici, segons és jutjat pels estàndards culturals de l'època. Podem donar ara una definició més completa dels trastorns mentals com a disfuncions perjudicials:

Una afecció és un trastorn mental si, i només si, (a) l'afecció causa algun mal o privació de benefici a la persona segons els estàndards de la cultura de la persona (el criteri de valor), i (b) l'afecció resulta de la incapacitat d'algun mecanisme mental per a realitzar la seua funció natural, tenint en compte que una funció natural és un efecte que forma part de l'explicació evolutiva de l'existència i estructura del mecanisme mental (el criteri explicatiu) (WAKEFIELD 1992, 385)

És important apreciar el rol que juguen totes dues parts de la definició, que es complementen mútuament: en primer lloc, restringir els trastorns a disfuncions en un sentit evolutiu evita que qualsevol afecció, comportament, resposta emocional, etc. que jutgem com a negativa es convertisca en un trastorn mental, la qual cosa és de gran importància donades les implicacions pràctiques que estan associades als trastorns mentals, com per exemple la necessitat de tractament. D'aquesta manera, per exemple, podrien evitar-se abusos com els ocorreguts en el cas de la drapetomania, *la malaltia que provoca que els negres fugen*, una suposada malaltia mental patida pels esclaus que, desafiant la seua naturalesa, decidien fugir i buscar la llibertat (CARTWRIGHT 1851). Per una altra banda, la condició que la disfunció resulte perjudicial o nociva per a l'individu també és important, perquè evita que qualsevol disfunció es convertisca automàticament en trastorn, de manera que, per dir-ho així, no ens vegem constrets pels fets biològics, sinó que deixem un espai per als valors morals i preferències socials. Així, per exemple, si finalment arribàrem a conèixer que l'homosexualitat involucra la disfunció d'algun mecanisme biològic, això no la convertiria necessàriament en un trastorn, perquè encara caldria valorar-ne l'impacte sobre l'individu.

Encara que la majoria de les crítiques a la proposta de Wakefield han estat dirigides a la primera part de la seua definició —com veurem tot seguit—, la veritat és que la segona part no està exempta de problemes. Tornem a reflexionar sobre el cas de l'homosexualitat: si arribara a conèixer-se que, efectivament, aquesta involucra algun tipus de disfunció biològica, com caldria aplicar-s'hi el segon criteri? Hauríem de definir-ne el caràcter perjudicial tan sols mirant els valors i preferències socials —el que, de fet, és socialment considerat com un perjudici— o hi hauria espai per a dur a terme una reflexió sobre el que hauria de ser considerat com a perjudicial, independentment de la seua consideració

social actual? És a dir, és possible que, de fet, l'homosexualitat estiga socialment considerada com a perjudicial perquè, per exemple, es considere que impedeix el ple desenvolupament de valors que se suposen centrals en la vida humana, com el de la família tradicional. En canvi, aquesta visió es pot confrontar argumentant sobre el que constitueix una vida o desenvolupament ple, la qual cosa no ha de passar necessàriament per la formació tradicional d'una família. No oblidem que, en última instància, aquells que revisen i elaboren els manuals diagnòstics són els qui han de decidir com aplicar aquest segon criteri: els experts poden reconèixer que, objectivament, certa afecció és entesa socialment com un perjudici o poden dur a terme una reflexió ètica per a valorar si una afecció suposa o no un perjudici. En el primer cas, malgrat que s'està tractant amb una valoració, ho estariem fent de manera objectiva, dedicant-nos a realitzar una investigació social, empírica, sobre la valoració que una societat fa de certa afecció. En el segon cas, el cas en què realitzarem una avaluació ètica al respecte, la nostra investigació seria objectiva o subjectiva depenent de si considerem que és possible mantindre una visió realista —que oferisca coneixement objectiu— sobre els valors morals o no.

Quant a la part referent a la disfunció, Wakefield també s'enfronta a crítiques importants, alguna més fàcilment salvable que d'altres. La primera d'elles, tal com la recull Kingma (2013), és la dels *trastorns seleccionats*, i apunta al fet que hi ha certs trets que foren seleccionats antigament però que hui dia considerem patològics pels efectes negatius que tenen en la nostra societat. Exemple d'això serien els comportaments antisocials, com les conductes extremadament violentes. Wakefield accepta aquesta observació i replica que en realitat no es problemàtica, ja que, efectivament, encara que hui dia considerem que això pot ser un trastorn mental, és merament perquè ho rebutgem i perquè considerem que pot beneficiar-se d'atenció mèdica.

Un altre problema, tal vegada de solució més complexa, és el dels *efectes no seleccionats*: “si una malaltia mental depèn d'estructures psicològiques que no tenen funció evolutiva, llavors no poden ser realment malalties mentals” (MURPHY 2006, 82). Per a exemplificar això, podem pensar en el cas de la capacitat humana de llegir (KINGMA 2013). Si aquesta habilitat fora una funció, llavors els efectes de la lectura haurien d'explicar per què els humans desenvolupem aquesta capacitat. Però sabem que la lectura és un fenomen molt recent, de manera que no pot haver sigut seleccionada com Wakefield pretén, sinó que probablement descansa en mecanismes que van ser seleccionats per altres efectes. Si la lectura no és una funció, llavors la dislèxia no pot ser un trastorn. Davant això, tenim tres eixides: la primera, considerar l'argument dels *efectes no seleccionats* com un argument fatal contra la teoria de Wakefield; la segona, acceptar que la dislèxia no és un trastorn, la qual cosa no ha d'implicar que no

hàgem de tractar-lo mèdicament; la tercera —la via per la qual opta Wakefield— considerar que, per a parlar de disfunció, és suficient que els mecanismes que possibiliten la capacitat de llegir —que sí que tenen una funció— no estiguen funcionant correctament, encara que aquesta capacitat no siga una funció. La resposta de Wakefield (1999) és que la dislèxia és un trastorn perquè, encara que la lectura no existeix com una funció seleccionada evolutivament i, per tant, la dislèxia no pot consistir en la disfunció d'aquesta, sí que hi subjau una disfunció d'algun mecanisme seleccionat per a altres propòsits. El problema és que sembla plausible que moltes de les nostres funcions mentals siguen com la lectura, és a dir, que facen ús de mecanismes que van ser seleccionats per a altres propòsits. Estaríem, per tant, davant una multitud de trets que manquen de funció i, en conseqüència, de la possibilitat de no funcionar correctament. No és evident fins a quin punt podríem donar explicacions com la de la dislèxia —si és que aquesta ens convenç— per a tots ells.

Altres crítiques s'han dirigit cap a les dificultats pràctiques que hui dia trobem a l'hora d'emprar el concepte de trastorn mental com a disfunció perjudicial, a causa dels pocs coneixements que tenim sobre la funció biològica en el sentit evolutiu (MURPHY 2006, 44). Si haguérem d'omplir les pàgines dels manuals psiquiàtrics únicament amb aquells trastorns que *sabem* que involucren la disfunció d'un mecanisme mental, estalviaríem, sens dubte, enormes quantitats de paper (KINGMA 2013). Per a evitar això, Wakefield en rebaixa les exigències; de moment, encara que no en tinguem el coneixement científic, podem conformar-nos amb la intuïció per a inferir que certs mecanismes van ser seleccionats evolutivament:

No és necessari conèixer els detalls de l'evolució o dels mecanismes interns per a saber, per exemple, que els casos típics de trastorns del pensament, drogodependència, trastorns de l'estat d'ànim, disfunció sexual, insomni, trastorns d'ansietat, trastorns de l'aprenentatge, etc., són fracassos d'alguns mecanismes per a realitzar les seues funcions dissenyades; això és obvi per les característiques de la superfície (WAKEFIELD 1997, 256)

Una altra possible solució —encara que indesitjable per a la majoria— seria acceptar el revisionisme que sembla implicar la teoria de la disfunció perjudicial i assumir que moltes de les categories diagnòstiques actuals no apunten en realitat a trastorns mentals. Això no ha de significar, *per se*, que la psiquiatria no pugui encarregar-se'n: senzillament, assumiríem que hi ha problemes que apareixen a causa de les enormes diferències entre l'entorn en el qual evolucionem i en el qual vivim actualment. Hauríem de preguntar-nos, llavors, com mereixen ser classificades aquestes afeccions i quin tipus

de tractament mereixen, si és que en mereixen algun. Podria argumentar-se que, malgrat que no són trastorns mentals, la psiquiatria té competència per a tractar-los i alleujar el sofriment de qui els pateix, o senzillament, millorar el funcionament normal humà. Segons el parer d'alguns, aquest moviment mostraria com és d'irrellevant el concepte de disfunció perjudicial, perquè a efectes pràctics tot seguiria igual: continuaríem tractant aquestes condicions en tant que impliquen perjudicis i riscos, només que ara els hauríem posat un altre nom (BOLTON 2007). Al nostre judici, per contra, el fet de posar-los un altre nom seria un pas important, perquè mostraria clarament que involucren preferències i valors en un sentit fort, la qual cosa té un important impacte en la seua legitimitat, així com en el seu tractament.

Siga com siga, malgrat totes les crítiques que la proposta del trastorn mental com a disfunció perjudicial ha rebut, la veritat és que actualment aquesta és la postura que gaudeix de major acceptació, la qual cosa es deu, segons el parer de molts, no al fet que siga especialment adequada, sinó al fet que no comptem amb cap proposta alternativa millor. Com ja avançàvem, el mateix Wakefield ha fet bon ús de la seua definició de trastorn mental per a fer activisme en contra de la inflació diagnòstica del Trastorn Depressiu Major, argumentant que els criteris diagnòstics que actualment recullen els manuals —tant el DSM-5 com el ICD-10, encara que especialment el primer— estan confonent el funcionament normal d'uns mecanismes mentals que van ser seleccionats per a superar la pèrdua (els quals poden produir una intensa tristesa, però sempre amb finalitats adaptatives) amb un mal funcionament de tals mecanismes, que en aquest cas s'haurien tornat disfuncionals i mancarien ja, per tant, de funció adaptativa produint, ara sí, una tristesa patològica o depressió.

#### PER A CONCLoure

El debat sobre la definició de la malaltia o trastorn mental continua sense estar resolt a hores d'ara. Encara que algunes de les preocupacions, tant dels teòrics de la medicina i la psiquiatria com de la societat en general, semblen haver-se desplaçat en les últimes dècades —de la malaltia mental greu als trastorns que difícilment poden ser distingits de la normalitat—, la veritat és que la qüestió nuclear sobre *de quina norma es desvia la malaltia mental* continua estant en el cor dels interrogants que ens acuiten.

Donada l'expansió de la psiquiatria en les últimes dècades i, sobretot, l'ingent augment del consum de psicofàrmacs, avançar en el debat és de crucial importància perquè té enormes repercussions pràctiques. Independentment que puguem arribar a conèixer de manera objectiva el vessant biològic dels

trastorns mentals —assumint que tal cosa existisca— el cert és que molt difícilment podem dubtar que la medicina —i encara més la psiquiatria— és molt més que el tractament d'entitats biològiques, perquè aquestes les pateixen persones, inserides en una cultura particular, amb una subjectivitat que, sens dubte, es veu impactada per l'experiència de la malaltia. La medicina, encara que pugua recolzar-se en la ciència de la biologia, no és merament ciència, sinó que és eminentment una pràctica social. Per a concloure aquest treball, ens agradaria portar a col·lació una observació que realitzen Ivana Marková i Germán Berrios sobre la psiquiatria contemporània:

La naturalesa 'anormal' de certs estats mentals i comportaments es determina sobre bases que són estudiades per les ciències humanes (no per les naturals), és a dir, en aquelles disciplines que tracten de comprendre les maneres en què els individus i les societats perceben, valoren i donen sentit al seu món en un moment donat. Una vegada que una conducta és demarcada com a anormal per prescripció social, les ciències naturals prenen el control i busquen associacions entre ella i els estats i funcions cerebrals. En algun moment d'aquest procés, allò anormal es converteix en patològic, és a dir, el seu significat és assumit per una correlació estadística que mostra una connexió entre el comportament i una localització cerebral (MARKOVÁ i BERRIOS 2012, 222)

Aquests paràgrafs contenen dues idees importants que voldríem remarcar: en primer lloc, mostren com —almenys fins ara— per tal que el vessant de la psiquiatria que descansa en les ciències naturals es pose en funcionament —buscant, per exemple, correlats en forma de neuroimatge a un símptoma mental— primer hem d'haver identificat aquest símptoma —certa conducta, emoció, pensament, etc.— com a anormal; és només llavors que podem anar a la recerca de la seua base biològica. En segon lloc, deixen entreveure que ací el problema de fons es troba en l'etern contrast entre *descripció* i *prescripció*; en aquest sentit, les ciències naturals només podrien descriure, mentre que la prescripció és terreny exclusiu de les ciències humanes. És per això que, encara que aconseguírem trobar aquelles descripcions tan desitjades en termes evolutius sobre la funció biològica que ens convenceren sobre quina era *la intenció* que tenia la naturalesa a l'hora de *dissenyar* cert mecanisme, així i tot, seguiríem en el terreny de la descripció i res s'hauria dit encara sobre *com ha de ser*. És més, segons la idea anterior, de fet només ens fixaríem en com funciona cert mecanisme en certes persones quan prèviament ja haguérem determinat que alguna cosa —el seu comportament— no estava funcionant *com ho hauria de fer*. La moralitat d'això és que, per lloables que siguem les intencions dels naturalistes que pretenen eliminar les qüestions de valor del si de la psiquiatria



per tal que no puguem tornar a cometre's abusos emparats sota la legitimitat mèdica, la veritat és que estem obrint la porta a perills molt similars, perquè de nou estem caient en el mateix error: ocultar les qüestions de valor que mouen la pràctica mèdica i psiquiàtrica, quan, per contra, hauríem d'esforçar-nos a traure a la llum les qüestions de valor que s'hi imbriquen per a poder reflexionar i deliberar degudament.

## BIBLIOGRAFÍA

- ANDREASEN, N. C. 1995, "The Validation of Psychiatric Diagnosis: New Models and Approaches", *American Journal of Psychiatry* 152 (2): 161-2.
- BALLESTEROS, V. 2018A, "¿Tristes o enfermos? De cómo la tristeza intensa se convirtió en patología", *Patologías de la existencia*, Saragossa: Servicio de Publicaciones de la Universidad de Zaragoza, 149-57.
- BALLESTEROS, V. 2018B, "The Metaphysics behind Pharmacotherapy: Treating Depression with Conventional and Psychedelic Drugs", *Psicología, Conocimiento y Sociedad* 8 (2): 7-28.
- BARKER, K. K. 2010, "The Social Construction of Illness: Medicalization and Contested Illness", *Handbook of Medical Sociology*, C. E. BIRD, P. CONRAD, A. M. FREMONT i S. TIMMERMANS (ed.), Nashville: Vanderbilt University Press, 6a ed., 147-62.
- BASAGLIA, F. 1967, *Che cos'è la psichiatria?*, Torí: Einaudi.
- BASAGLIA, F. 1968, *L'istituzione negata*, Torí: Einaudi.
- BENTALL, R. P. 2009, *Doctoring the mind: is our treatment of mental illness really any good?*, Nova York: New York University Press.
- BERRIOS, G. E. 2011, *Hacia una nueva epistemología de la psiquiatría*, A. GIACCONE i M. V. SCHMIDT (trad.), Buenos Aires: Polemos.
- BOLTON, D. 2007, "The Usefulness of Wakefield's Definition for the Diagnostic Manuals", *World Psychiatry* 6 (3): 164-5.
- BOORSE, C. 1975, "On the Distinction between Disease and Illness", *Philosophy & Public Affairs* 5 (1): 49-68.
- BOORSE, C. 1976. "What a Theory of Mental Health Should Be", *Journal for the Theory of Social Behaviour* 6 (1): 61-84.
- BOORSE, C. 1977, "Health as a Theoretical Concept", *Philosophy of Science* 44 (4): 542-73.
- BROWN, P. 1995, "Naming and Framing: The Social Construction of Diagnosis and Illness", *Journal of Health and Social Behavior* 35: 34-52.
- BUTLER, J. 2011, *Bodies That Matter: On the Discursive Limits of "Sex"*, Routledge Classics. Abingdon, Oxon ; Nova York: Routledge.
- CANGUILHEM, G. 2011, *Lo normal y lo patológico*, Mèxic: Siglo Veintiuno Editores.
- CARTWRIGHT, S. 1851, "Diseases and Peculiarities of the Negro Race", *De Bow's Review* XI. <http://www.pbs.org/wgbh/aia/part4/4h3106t.html>.
- CONRAD, P. i BARKER, K. K. 2010, "The Social Construction of Illness: Key Insights and Policy Implications", *Journal of Health and Social Behavior* 51 (1\_suppl): S67-79.

- DRESCHER, J. 2015, "Out of DSM: Depathologizing Homosexuality", *Behavioral Sciences* 5 (4): 565-5.
- EISENBERG, L. 1988, "The Social Construction of Mental Illness", *Psychological Medicine* 18 (1): 1-9.
- EISENBERG, L. 2007, "Commentary with a Historical Perspective by a Child Psychiatrist: When 'ADHD' Was the 'Brain-Damaged Child'", *Journal of Child and Adolescent Psychopharmacology* 17 (3): 279-83.
- ENGELHARDT, H. T. 1976, "Ideology and Etiology", *Journal of Medicine and Philosophy* 1 (3): 256-68.
- ERESHEFSKY, M. 2009, "Defining 'Health' and 'Disease'", *Studies in History and Philosophy of Biological and Biomedical Sciences* 40 (3): 221-7.
- EYSENCK, H. J. 1960, "Classification and the Problem of Diagnosis", *Handbook of Abnormal Psychology: an Experimental Approach*, H. J. EYSENCK (ed.), Londres: Pitman Medical Publishing, 1-31.
- FOUCAULT, M. 1963, *Naissance de la clinique: une archéologie du regard médical*, Paris: Presses Universitaires de France.
- FOUCAULT, M. 1966, *Les mots et les choses: une archéologie des sciences humaines*, Paris: Gallimard.
- FRANCES, A. 2014, *Saving Normal: An Insider's Revolt against out-of-Control Psychiatric Diagnosis, DSM-5, Big Pharma, and the Medicalization of Ordinary Life*.
- FREIDSON, E. 1970, *Profession of Medicine: A Study of the Sociology of Applied Knowledge*, Nova York: Dodd, Mead & Company.
- FULFORD, K. W. M. 2001, "'What Is (Mental) Disease?': An Open Letter to Christopher Boorse", *Journal of Medical Ethics* 27 (2): 80-5.
- FULFORD, K. W. M. i VAN STADEN, C. W. 2013, "Values-Based Practice: Topsy-Turvy Take-Home Messages from Ordinary Language Philosophy (and a Few Next Steps)", *The Oxford Handbook of Philosophy and Psychiatry*, K. W. M. FULFORD, M. DAVIES, R. G. T. GIPPS, G. GRAHAM, J. Z. SADLER, G. STANGHELLINI i T. THORNTON (ed.), International Perspectives in Philosophy and Psychiatry, Oxford: Oxford University Press, 385-410.
- GIROUX, É. 2010, *Après Canguilhem: définir la santé et la maladie*, Paris: Presses Universitaires de France.
- HACKING, I. 1996, "The Looping Effects of Human Kinds", *Causal Cognition*, D. SPERBER, D. PREMACK i A. J. PREMACK (ed.), Oxford University Press, 351-83.
- HACKING, I. 1999, *The Social Construction of What?*, Cambridge, Mass: Harvard University Press.
- HASLANGER, S. A. 2012, "Ontology and Social Construction", *Resisting Reality: Social Construction and Social Critique*, Nova York: Oxford University Press, 83-112.
- HERNÁEZ, A. M. 2011, "La copia de los hechos. La biomedicina, el poder y sus encubrimientos", *Quaderns* 27: 45-64.

- HOFMANN, B. 2002, "On the Triad Disease, Illness and Sickness", *The Journal of Medicine and Philosophy* 27 (6): 651-73.
- HORWITZ, A. V. i WAKEFIELD, J. C. 2007, *The loss of sadness: how psychiatry transformed normal sorrow into depressive disorder*, Oxford ; Nova York: Oxford University Press.
- INSEL, T.; CUTHBERT, B.; GARVEY, M.; HEINSEN, R.; PINE, D. S.; QUINN, K.; SANISLOW, C. i WANG, P. 2010, "Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders", *American Journal of Psychiatry* 167 (7): 748-51.
- KENDELL, R. E. 1975, "The Concept of Disease and Its Implications for Psychiatry", *The British Journal of Psychiatry* 127 (4): 305-15.
- KENDELL, R. E. 1986, "What are mental disorders?", *Issues in psychiatric classification: Science, practice and social policy*, Nova York: Human Sciences Press, 23-45.
- KINGMA, E. 2007, "What Is It to Be Healthy?", *Analysis* 67 (294): 128-33.
- KINGMA, E. 2013, "Naturalists Accounts of Mental Disorder", *The Oxford Handbook of Philosophy and Psychiatry*, K. W. M. FULFORD, M. DAVIES, R. G. T. GIPPS, G. GRAHAM, J. Z. SADLER, G. STANGHELLINI i T. THORNTON (ed.), International Perspectives in Philosophy and Psychiatry, Oxford: Oxford University Press, 363-84.
- KIRSCH, I. 2010, *The emperor's new drugs: exploding the antidepressant myth*, Nova York: Basic Books.
- KRAMER, P. D. 1993, *Listening to Prozac*, Nova York: Viking.
- LAING, R. D. 1960, *The Divided Self: An Existential Study in Sanity and Madness*, Londres: Tavistock Publications.
- LAING, R. D. 1967, *The politics of experience and The bird of paradise*, Harmondsworth: Penguin Books.
- LANE, C. 2007, *Shyness: how normal behavior became a sickness*, New Haven: Yale University Press.
- MARGOLIS, J. 1959, "Illness and Medical Values", *The Philosophy Forum* 8: 55-76.
- MARKOVÁ, I. S. i BERRIOS, G. E. 2012, "Epistemology of Psychiatry", *Psychopathology* 45 (4): 220-7.
- MELLOR MARSÁ, B. i ARAGONA, M. 2009, "De la categoría a la dimensión: una mirada crítica a la evolución de la nosografía psiquiátrica", *Revista de la Asociación Española de Neuropsiquiatría* 29 (103): 217-28.
- MURPHY, D. 2006, *Psychiatry in the scientific image*, Philosophical psychopathology, Cambridge, Mass: MIT Press.
- NORDENFELT, L. 1993, "Concepts of Health and Their Consequences for Health Care", *Theoretical Medicine* 14 (4): 277-85.
- NORDENFELT, L. 2007, "The Concepts of Health and Illness Revisited", *Medicine, Health Care and Philosophy* 10 (1): 5-10.

- PARSONS, T. 2005, *The Social System*, Taylor & Francis E-Library. Londres: Routledge.
- PUTNAM, H. 2002, *The Collapse of the Fact/Value Dichotomy*, Cambridge, Mass.: Harvard University Press.
- SCADDING, J. G. 1967, "Diagnosis: The Clinician and the Computer", *The Lancet* 21: 877-82.
- SEARLE, J. R. 1995, "Does the real world exist? Part I: Attacks on realism", *The Construction of Social Reality*, Nova York: The Free Press, 149-76.
- SEDGWICK, P. 1982, *Psychopolitics*, Londres: Pluto Press.
- SHORTER, E. 1997, *A history of psychiatry: from the era of the asylum to the age of Prozac*, Nova York: John Wiley & Sons.
- SPITZER, R. L.; ENDICOTT, J. i MICOULAUD FRANCHI, J.-A. 2018, "Medical and Mental Disorder: Proposed Definition and Criteria", *Annales Médico-Psychologiques, Revue Psychiatrique* 176 (7): 656-65.
- SZASZ, T. 1972, *The myth of mental illness: foundations of a theory of personal conduct*, New and abbreviated ed., Londres: Paladin.
- SZASZ, T. 1973, *Ideology and insanity: essays on the psychiatric dehumanization of man*, Londres: Calder and Boyars.
- SZASZ, T. 1984, *The therapeutic state: psychiatry in the mirror of current events*, Buffalo, N.Y: Prometheus Books.
- TIMMERMANS, S. i HAAS, S. 2008, "Towards a Sociology of Disease", *Sociology of Health & Illness* 30 (5): 659-76.
- TURNER, B. S. 1995, *Medical power and social knowledge*, 2a ed., Londres: Sage Publications.
- TURNER, B. S. 2008, *The Body & Society: Explorations in Social Theory*, 3a ed., Londres: Sage Publications.
- WAKEFIELD, J. C. 1992, "The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values", *American Psychologist* 47 (3): 373-88.
- WAKEFIELD, J. C. 1997, "Normal Inability Versus Pathological Disability: Why Ossorio's Definition of Mental Disorder Is Not Sufficient", *Clinical Psychology: Science and Practice* 4 (3): 249-58.
- WAKEFIELD, J. C. 1999, "Evolutionary versus prototype analyses of the concept of disorder", *Journal of Abnormal Psychology* 108 (3): 374-99.
- WAKEFIELD, J. C. 2015, *Sadness or Depression? International Perspectives on the Depression Epidemic and Its Meaning*, Nova York: Springer Berlin Heidelberg.
- WHITAKER, R. 2009, "Deinstitutionalization and Neuroleptics: The Myth and the Reality", *Psychotherapeutic Approaches to Schizophrenic Psychoses: Past, Present and Future*, Y. O. ALANEN, M. GONZÁLEZ DE CHÁVEZ, A.-L. S. SILVER i B. MARTINDALE (ed.), London, Nova York: Routledge, 346-56.

- WHITAKER, R. i COSGROVE, L. 2015, *Psychiatry Under the Influence: Institutional Corruption, Social Injury, and Prescriptions for Reform*, Nova York: Palgrave Macmillan US : Imprint : Palgrave Macmillan.
- WILLIAMS, S. J. 2005, "Parsons Revisited: From the Sick Role To...?", *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine* 9 (2): 123-44.
- WILLIAMS, S. J. 2006, "Medical Sociology and the Biological Body: Where Are We Now and Where Do We Go from Here?", *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine* 10 (1): 5-30.

# *Ressenyés*





Julien CANAVERA 2018, *Pensar problemàticament. Ensayo sobre Gilles Deleuze*, València: Institució Alfons el Magnànim. ISBN 978-84-7822-742-6, 380 pàgines.

Foucault va dir que potser un dia el segle seria deleuzià. Aqueixa advertència de Foucault pren un regust premonitori a la vista de les societats en què vivim, les quals s'assemblen cada vegada més a les que ja des dels anys 70 Deleuze anticipava amb Guattari. Però si el segle està esdevenint deleuzià no ho és tant perquè es puga trobar en Deleuze una mena de *telos* de la filosofia, ni tan sols que arribe a ser l'hegemònica, sinó més bé perquè se'ns fa impossible comprendre les nostres societats sense fer-nos càrrec de tota una sèrie de problemàtiques que podríem anomenar "deleuzianes": la relació entre el poder i el desig, el paper de la informació en les societats digitals o l'esfondrament de part de les institucions que havien regit la vida durant la Modernitat (la família nuclear i el matrimoni heterosexual, la fàbrica com a centre de treball i el treball material com a principal font de valor, etc.), alhora que n'han aparegut altres de noves que han transformat la manera com existim i ens relacionem (l'aparició de les xarxes socials, la transformació i deslocalització del treball fabril, la hiperconnectivitat a què estem sotmesos, etc.).

Ara bé, la filosofia política de Deleuze no es comprèn des de les teories modernes del subjecte, l'acció i la intencionalitat, sinó que s'assenta sobre el que l'autor anomena empirisme transcendent: un empirisme (ja que parteix de l'experiència) en què ja no és el subjecte, sinó l'experiència real mateixa la que imposa l'ordre i el sentit.

Aquest text, que ha publicat recentment la Institució Alfons el Magnànim, ens serveix per accedir al pensament d'un dels filòsofs del segle xx que més vigència conserva i que esdevé fonamental quan tractem de comprendre el pensament polític a partir del Maig del 68. *Pensar problemàticament. Ensayo sobre Gilles Deleuze*, signat per Julien Canavera, es presenta com una introducció a l'obra del filòsof francès que aspira a donar una imatge general del seu pensament, recorrent les seues "múltiples entrades i galeries", i articulant els distints problemes presents en el pensament deleuzià.

El treball de Canavera, força documentat, consisteix a desplegar el pla d'immanència deleuzià a partir del concepte de "vida", present a Nietzsche i Bergson, tal i com el reprèn l'autor a *Différence et répétition*, una immanència de si, mai tancada, que es troba en permanent esdevenir, i l'exposa a partir de

quatre temes principals. En primer terme, la vida com a pla d'immanència és la instància última a partir de la qual interpretar i valorar, ja que enclou la totalitat del sentit, allunyant-se de les teories essencialistes hereves del platonisme; en segon terme, trobem una exposició de la imatge del pensament deleuziana que fa incís en la crítica al pensament natural i la intencionalitat, tractant de mostrar el caràcter triple del pensament: exterior, relacional i presubjectiu. Des d'ací és un pas el que hi ha per presentar el subjecte com a interpretació i com a procés productiu; és el que Canavera fa al tercer capítol, presentant un subjecte a la deriva com una instància constantment arrossegada, d'una banda a altra, per les forces que el situen, l'ubiquen i el fan fulgurar com a resultat: un punt de vista que sorgeix del Defora com a interpretació. Finalment s'exploren les temptatives de Bergson i Nietzsche per traçar un nou concepte d'esdeveniment i les conseqüències directes que presenta per a l'ontologia aquesta concepció de la temporalitat del Ser que confronta amb qualsevol temptativa de pensar-lo com si fos donat tot d'una, complet i tancat, idèntic a si mateix, mitjançant el qual Deleuze tracta de desempallegar-se de concepcions normatives de la filosofia, tot arribant a un lloc estretament vinculat amb el que proposava Nietzsche per convertir la vida mateixa en una obra d'art.

Aquesta presència de Nietzsche i Bergson és, a més a més, una constant on retorna l'assaig per remarcar aquesta concepció temporal del Ser de què Deleuze és deutor, i que li permet fugir d'aquesta concepció representativa, incapaç de donar compte de la vida, per posar el focus d'atenció en l'esdevenir, que es converteix, en aquest vitalisme deleuzià que Canavera presenta, en una instància central des de la qual es valora i d'acord amb la qual s'avalua.

És aquest concepte de vida el que permet a Canavera fer-se càrrec de la major part dels problemes que aborda l'obra de Deleuze. La transcendència instaurada pel platonisme i la seua inversió, la crítica a la identitat última del Ser amb el que és i a la imatge dogmàtica del pensament, la mort —però encara no del tot— del subjecte, subjecte que està sempre en crisi, que sempre deixa de ser ell, que ara ja és només desig, i el desig mateix, únic motor de la vida i el pensament. Així Canavera presenta i proposa eixides al projecte de Deleuze d'invertir el platonisme i de produir una nova imatge del pensament.

Al respecte del problema de la subjectivitat en el pensament deleuzià, Canavera mostra, servint-se principalment de *Différence et répétition* (1968) que la qüestió de la mort del subjecte, tesi que pren força a partir de *L'Anti-Èdipe* (1972), i més encara a *Mille Plateaux* (1980), ja està present en els primers lineaments de la seua filosofia. No obstant això, trobem en l'aproximació que fa Canavera al pensament de Deleuze una revaluació de la institució. Encara que el subjecte ocupa un lloc de terminal en el procés productiu (el que podríem anomenar activitat genèrica o simplement "vida"), aquest és una

instància sempre present, que sempre sorgeix amb la interpretació, si bé com a subproducte d'aquesta, en un lloc "actiu" de l'operació del pensament, pel que no podem descuidar-nos d'aquests processos que conformen la manera com ens relacionem amb el món. És per això que es torna una tasca necessària la transformació dels dispositius de producció de subjectivitat i les tècniques de govern, articulant noves formes de relacions de poder que permeten la formació de subjectivitats més aptes per a la vida.

Tot aquest problema ve reconstruït per Canavera a partir de les lectures que fa Deleuze de Hume, Kant, Spinoza, Nietzsche i Simondon, principalment, per mostrar els processos d'individuació que es donen sota la màscara del subjecte. D'altra banda, l'escassa presència de la psicoanàlisi en el text, que queda relegada a un lloc perifèric —però compensada amb una àmplia exposició de l'ontologia i de la subjectivitat—, ens permet preguntar-nos si, com el propi Deleuze diu amb Guattari al respecte de Kafka, la psicoanàlisi és una influència determinant per al seu pensament o bé només l'empra com quelcom que té al seu abast, que forma part del pensament de la seua època i en relació al qual pot expressar les seues idees, sense quedar determinat per una influència estreta seua ni veure's obligat a retre-li comptes.

A banda de totes aquestes qüestions que hem considerat centrals, el text de Canavera aborda amb exhaustivitat altres qüestions que de primeres pot semblar que ocupen un segon ordre d'importància però que és gràcies a elles que pot elaborar una cartografia del pensament deleuzià que, lluny de mostrar una solució fàcil i autocomplaent dels problemes que planteja, fa incís en la complexitat i la profunditat que mostren. Aquesta voluntat de fugir de tesis simplistes i reduccionistes del pensament de Deleuze converteixen l'estudi de Canavera en un material sens dubte a tenir en compte en tasques de recerca per accedir amb pas ferm al pensament del primer Deleuze —és a dir, el dels monogràfics d'història de la filosofia i de les seues dues grans obres anteriors a la col·laboració amb Guattari, que s'estén des de 1972 a 1993.

Una particularitat que caldria ací també ressenyar és l'excel·lent domini de la llengua castellana que Canavera mostra en aquest assaig de prosa barroca que, val a dir, de vegades es converteix en una fulla de doble fil per a lectors no especialitzats, ja que de vegades no és senzill seguir el ritme proposat. No obstant això —i gràcies, en part, a això—, tot i l'atenció que arriba a exigir, el text està sòlidament articulat; de fet, entre les principals virtuts del text cal remarcar aquesta harmonia que troben els distints temes dintre de l'assaig, entrelaçats de manera acurada i rigorosa, respectant a *la lettre* els textos deleuzians, que estan ordits ací amb enginy i desimboltura, així com el gran rizoma que produeix en l'interior de l'obra del parisenc, que queda ací oberta als comentaris i travessada per les interpretacions que s'han fet al respecte.

Sense cap mena de dubte, es tracta d'un material interessant que pot ser útil en diferents registres, per una banda com a introducció general a l'autor, però, i no en menor mesura, també com una referència a tenir en compte per als estudis especialitzats, gràcies a la solvència amb la qual presenta el pensament de Deleuze i el domini mostrat de les fonts.

RAIMON RIBERA AÑÓ  
*Universitat de València*