Published in 2012 in *International Studies in the Philosophy of Science.*

Sabina Leonelli

ESRC Centre for Genomics in Society, University of Exeter

EX4 4PJ Exeter, UK

Tel. 0044-1392-725140

Fax 0044-1392-724676

s.leonelli@exeter.ac.uk

Classificatory Theory in Data-Intensive Science: The Case of Open Biomedical Ontologies

Abstract

Knowledge-making practices in biology are being strongly affected by the availability of data on an unprecedented scale, the insistence on systemic approaches and growing reliance on bioinformatics and digital infrastructures. What role does theory play within data-intensive science, and what does that tell us about scientific theories in general? To answer these questions, I focus on the Open Biomedical Ontologies, digital classification tools that have become crucial to sharing results across research contexts in the biological and biomedical sciences, and argue that they constitute an example of classificatory theory. This form of theorising emerges from classification practices in conjunction with experimental know-how and expresses the knowledge underpinning the analysis and interpretation of data disseminated online.

Keywords: bio-ontologies; theory; data; data-intensive science; classification; databases; evidence; bioinformatics.

1. Introduction: Questioning the Role of Theory in Data-Intensive Science

Over the last three decades, new ways to disseminate and analyse scientific data have facilitated a large shift in research practices. This is not simply a quantitative shift, brought about by the speed and quantity of data available online: it is a qualitative shift in how scientific research is carried out, with important consequences for what counts as scientific knowledge and how that knowledge is obtained and used. Prominent scientists have characterised this shift as leading to a new, 'data-intensive' paradigm for research, encompassing innovative ways to produce, store, disseminate and interpret huge masses of data across several fields ranging from physics to climate science (Hey, Tansley and Tolle 2009). The biological sciences are one of the most fertile and receptive areas to these developments (Leonelli 2012). Thanks to high-throughput technologies for data production such as sequencing and microarray experiments, the collection of data within experimental biology has become increasingly fast and automated, resulting in the production of billions of data-points in need of a biological interpretation. Massive research efforts are devoted to the dissemination and modelling of data through the internet, by means of infrastructures such as online databases, in the hope that digital access will enable researchers to use these data to generate biological insights. Biologists can now gather and integrate data obtained on a wide variety of organisms by laboratories across the globe, no matter which specific expertises and interests guide the production of data at each of those locations (Rhee and Crosby 2005, Blake and Bult 2005, Leonelli 2009). The practices devoted to extracting inferences from data *in silico* are becoming increasingly sophisticated, resulting in discoveries obtained through the analysis of datasets available online (thus without carrying out experiments and/or data collection *in vivo*; see Buetow 2005). Examples of data-intensive discovery include:

- *corroborating a claim through the triangulation of evidence acquired on the same phenomenon through independently conducted inquiries*;[1] for instance, the discovery that

[1] For a sophisticated analysis of the notion of triangulation, see Wylie 2002.

the same genes or pathways play similar regulatory roles across species is often due to the possibility to retrieve and compare data that would have otherwise been buried in specific laboratories or within circumscribed disciplinary circles;

- *identifying new patterns or correlations through data mining*: a striking instance of this are so-called 'random walks' through data, where software is used to search existing datasets for statistically significant patterns (e.g. gene-pair interactions in biological networks of specific model organisms; Chipman and Singh 2009);
- *spotting gaps in the existing knowledge about a given entity*: the accumulation of evidence can point researchers towards areas of investigation that are not yet charted, as when discovering 'ultra-conserved regions' in vertebrate DNA and their regulatory role in development (Blake and Bult 2005).[2]

The examples above capture new ways in which computational and experimental practices are combined in order to make sense of existing data. In all of these cases, computational tools for data analysis are assigned a prominent role in facilitating the extraction of patterns from data, while experimental work is conceived as means to verify and explain those patterns. Indeed, data-intensive research as a whole could be defined as fostering the use of automated inferential procedures from data available online as a starting point for inquiry.

This situation raises a deep philosophical question, namely how to characterise the methodology and epistemic significance of these practices, and particularly the role of theory and hypotheses in this context. This question is hotly debated within scientific circles, where proponents of this approach argue that *in silico* analyses of high-throughput datasets are giving rise to a new kind of

[2] Of course, there have long been expectations in molecular systematics that there would be highly conserved areas of the genome, and that these will serve to indicate relatedness. Data-intensive methods offer new, efficient ways to spot these areas, thus making it possible to explore their significance.

epistemology, one in which the automated analysis of evidence claims primacy over traditional practices of experimentation, theorisation and hypothesis-testing (Hey, Tansley and Tolle 2009). This paper aims to explore this claim from a philosophical perspective, by exploring the role played by theories in the design, development and use of computational tools for data analysis. My analysis stands in contrast to a simplistic understanding of data-intensive science as juxtaposed to 'hypothesis-driven' research, i.e. as based on the inductive inference of patterns from datasets leading to the formulation of testable claims without recourse to pre-conceived hypotheses. As recognised by both champions and critics of data-intensive research (Allen 2001, Kell and Oliver 2004, O'Malley et al 2009), extracting biologically meaningful inferences from data involves a complex interplay of components, such as reliance on background knowledge, model-based reasoning and iterative tinkering with materials and instruments. Thus, a simplistic opposition between inductive and deductive procedures does not help to understand the epistemic characteristics of this research mode. The question that needs to be asked about data-intensive science is not whether it includes some form of theoretical assumption, which it certainly does, but rather whether it uses theories in a way that distinguishes it from other forms of inquiry – and what does this tell us about the epistemology of current research.

The scientific context within which I address this question is model organism biology, which encompasses several instances of data-intensive science given its increasing reliance on databases for the storage, retrieval, analysis and modelling of data. Research on well-established model organisms (such as the plant *Arabidopsis thaliana*, the worm *Caenorhabditis elegans*, the zebrafish *Danio rerio* and the fruit-fly *Drosophila melanogaster*) has grown hand in hand with databases and repositories that freely and openly disseminate data across research communities around the globe (Bult 2006, Leonelli and Ankeny 2012). Within this context, a classification system was developed that is proving crucial to the ordering, analysis and re-use of data for purposes of discovery: the *bio-ontology*. Bio-ontologies enable researchers from different disciplines and locations to share resources of common interest through the Internet, and use them to further research (Ashburner et al 2000, Backlawski and Niu 2006). For the purposes of this paper, I restrict my examination to the bio-ontologies collected by the Open Biomedical Ontologies [OBO] Consortium, an organisation founded to facilitate communication and coherence among bio-ontologies with broadly similar characteristics (Smith et al 2007). In particular, I shall focus on one of the bio-ontologies in the OBO Consortium, the Gene Ontology, which is widely regarded as the most successful case of bio-ontology construction to date and used as a template for developing several other prominent bio-ontologies (Ashburner et al 2000, Brazma et al 2006). It is important to note the specific focus of my discussion at the outset, because since the development of the Gene Ontology several other formalisations for bio-ontologies were introduced and the status and place of the OBO format among them is hotly disputed (Egaña Aranguren et al 2007). For instance, ontologies constructed using the Web Ontology Language (OWL) are increasingly attracting attention as a useful alternative to the OBO format, and efforts are under way to make the two systems compatible (e.g. Hoehndorf 2010; Hamid Tirmizi et al 2011). These differences and convergences are crucial to the future development of bio-ontologies and the extent of their popularity in model organism research; however, they are not relevant to my philosophical analysis, which focuses on the features currently adopted by the OBO Consortium *in order to serve the community of experimental biologists that uses them*. The Gene Ontology is particularly relevant to my argument not only because of its foundational status among bio-ontologies, but also because it has been explicitly developed in order to facilitate data integration, comparison and re-use across model organism communities (Leonelli 2010). Bio-ontologies may be used for other endeavors, such as for instance managing and accessing data in the first place (that is, even before they are circulated beyond the laboratory where they are originally produced).[3] The Gene Ontology, as many within the OBO Consortium, is specifically devoted to represent the biological knowledge underlying the re-use of data within new

[3] A useful overview of the various purposes for which bio-ontologies are developed can be found in Keet (2010). Thanks to an anonymous referee for pointing this paper out to me.

research contexts: in other words, it defines the ontology that researchers need to share to successfully draw new inferences from existing datasets (Ashburner et al 2000, Lewis 2004, Renear and Palmer 2009). At the same time, the Gene Ontology is constantly modified depending on the state of research and the interests of their users, and the mechanisms through which it is updated make this bio-ontology particularly helpful in disseminating data for the purpose of discovery (Leonelli et al 2011). In this paper, I argue that understanding how bio-ontologies such as the Gene Ontology work is crucial in order to uncover the epistemic structure and modus operandum of data-intensive science, and particularly the role of theory within it.

The structure of the paper is as follows. After a brief overview of what bio-ontologies such as the Gene Ontology consist of and how they are developed, I argue that they exemplify the role played in data-intensive science by one specific form of theory, which I call *classificatory theory*. This way of theorising is strongly grounded in the interrelation between practices of description and classification on the one hand, and experimental practices on the other. I suggest that the roots of this type of theorising are deep, yet received relatively little attention in philosophy until now**;** and I discuss the significance of this analysis towards existing philosophical conceptions of classification and theory, as well as towards understanding data-intensive science.

1. Bio-ontologies: Representing Knowledge to Disseminate Data

Open Biomedical Ontologies are classification tools that have been recently developed to facilitate the dissemination of data across research contexts through digital databases. They are crucial to the functioning of databases in several ways: they provide a means to classify, retrieve and analyse vast amounts of data of different types; they store and organise those data on the basis of users' own research interests, which facilitates data retrieval; and they make data travel across research contexts, model organism communities and separate disciplines (Ashburner et al 2000; Smith et al 2007; Leonelli 2010). They consist of a network of related terms, where each term denotes a specific biological phenomenon and is used as a category to classify data relevant to the study of that phenomenon. In the words of two of the scientists who contributed to their development, bio-ontologies are ‘formal representations of areas of knowledge in which the essential terms are combined with structuring rules that describe the relationship between the terms. Knowledge that is structured in a bio-ontology can then be linked to the molecular databases’ (Bard and Rhee 2004).

Here is how this type of bio-ontologies work. Every term is assigned a precise definition, which describes the characteristics of the phenomenon which the term is intended to designate, sometimes including species-specific exceptions (Baclawski and Niu 2006, 35). For instance, the term ‘nucleus’ is defined as follows:

> ‘A membrane-bound organelle of eukaryotic cells in which chromosomes are housed and replicated. In most cells, the nucleus contains all of the cell’s chromosomes except the organellar chromosomes, and is the site of RNA synthesis andprocessing. In some species, or in specialized cell types, RNA metabolism or DNA replication may be absent’ (GO website, accessed July 2010).

Each bio-ontology term can be linked to a dataset through a process called *annotation*. This process, carried out by bio-ontology curators, includes searching biological publications for data that have reliably been associated with a given phenomenon, and that can therefore be safely assumed to provide evidence about that phenomenon. These data, which might have widely diverse provenance, are then classified together under the label provided by the bio-ontology term. As a result of annotation, biologists can type the terms that best describe their research interests into the search engine of any database that uses the bio-ontology, and get instant access to the data available on the phenomena of interest. The definitions assigned to bio-ontology terms are thus intended to be descriptive of phenomena – in other words, to capture existing knowledge about the features and components of actual biological entities and processes. At the same time, terms in bio-ontologies function as classificatory categories through which datasets can be organised and retrieved.

Terms are related through a network structure. The basic relationship between terms is called containment and it involves a *parent* term and a *child* term. The child term is *contained* by the parent term when the child term represents a more specific category of the parent term. This relationship is fundamental to the organisation of the bio-ontology network, as it supports a hierarchical ordering of the terms used. The criteria used to order terms are chosen in relation to the characteristics of the phenomena captured within each bio-ontology. For example, the Gene Ontology uses three types of relations among terms: 'is_a', 'part_of' and '--regulates'.[4] The first category denotes relations of identity, as in 'the nuclear membrane is a membrane'; the second category denotes mereological relations, such as 'the membrane is part of the cell'; and the third category signals regulatory roles, as in 'induction of apoptosis regulates apoptosis'. In other bio-ontologies, the categories of relations available can be more numerous and complex: for instance, including relations signalling measurement ('measured_as') or belonging ('of_a'). One way to visualise bio-ontologies is to focus on their hierarchical structure as a network of terms, as illustrated in figure 1. This approach is the one preferred by experimental scientists, who use it to focus on the ways in which terms are related to each other.

[figure 1]

Another, more philosophically interesting way to view bio-ontologies is to conceive of them as a series of descriptive propositions about biological entities and processes. So in the example

[4] Note that 'regulates' is not organized in the parent/child structure. Also, the Gene Ontology is now in the process of incorporating another two types of relations, 'has_part' and 'occurs_in', and could potentially adopt many other types of relations, as documented by Smith et al (2005).

provided in figure 1, the bio-ontology tells us that ‘cell development is part of cell differentiation’ and ‘cell development is a cellular process’. One can assign meaning to these statements by appealing to the definitions given to the terms ‘cell development’, ‘cell differentiation’ and ‘cellular process’, as well as to the relations ‘is_a’ and ‘part_of’. Similarly, one learns that ‘neural crest formation is part of neural crest cell development’ and that ‘neural crest formation is an epithelial to mesenchrymal transition’. Viewed in this way, bio-ontologies consist of a series of claims about phenomena: a text outlining what is known about the entities and processes targeted by the bio-ontology. As in the case of any other text, the interpretation given to these claims depends partly on the interpretation given to the definitions assigned to the terms and relations used; and the hierarchical structure given to the terms implies that changes to the definition of one term, or to the relationship linking two terms in the network, have the potential to shift the meaning of all other claims in the same network.

Yet, definitions are not the only tool available to researchers in order to interpret claims contained in a bio-ontology. Claims are also assessed through the evaluation of the data that has been linked to each of the terms involved. This is possible through the consultation of *meta-data*, i.e. information about the experimental context in which data were originally produced, which helps users to assess their evidential value for the purposes of their own research interests. Examples of meta-data include the specification of the organism(s) on which data were obtained; the names of researchers involved in that experiment and the publications through which it was disclosed; and the instruments, experimental set-up and protocols used to collect data (Brooksbank and Quakenbush 2006). The choice and insertion of meta-data is a crucial step in the development of a bio-ontology, because it enables users to investigate the experimental circumstances in which data retrieved through the bio-ontologies have been gathered, thus putting them in a position to evaluate the interpretation given to the data by their producers (if any) and the way in which they support (or not) the claims made in the bio-ontology. For example, a researcher interested in neural crest formation can investigate the validity of the claim ‘neural crest formation is an epithelial to mesenchymal transition’ by checking which data were used to validate that assertion (i.e. to link the term ‘neural crest formation’ with the term ‘epithelial to mesenchymal transition’ via the relation ‘is_a’), what organism they were extracted from, which procedures were used to obtain them and who conducted that research. As a result, the researcher can appeal to her own understanding of what constitutes a reliable experimental set-up, a trustworthy research group and an adequate model organism in order to interpret the quality of the data in question as evidence for the claim.

2. From Dissemination to Discovery

The opportunity for researchers to give empirical as well as conceptual meaning to claims about phenomena is what ultimately makes bio-ontologies different from a textbook and from other kinds of classification systems. Bio-ontologies like the Gene Ontology aim to incorporate two kinds of biological knowledge, both of which turn out to be indispensable to mining data for the purposes of discovery. The first is *propositional knowledge about phenomena*, which is represented as a series of interrelated claims as discussed above; the second is *practical knowledge about data production*, which is represented through meta-data. Much effort is invested in making sure that the choice of terms, definitions and relations represented in a bio-ontology is corroborated by existing, peer-reviewed literature and intelligible to users coming from different disciplines. The same is true of the choice of what counts as meta-data. Terms, definitions, relations and meta-data are selected and defined through consensus-seeking mechanisms implemented on a global scale, such as consultations and workshops. Knowledge represented in a bio-ontology has been carefully processed so as to be intelligible across as many contexts as possible (Leonelli 2010).

It is curators who are mainly responsible for processing existing knowledge for incorporation into bio-ontologies (Howe et al 2008). They formulate a definition for each term by consulting the available literature and experts in relevant fields. At the same time, they define the experimental provenance of data linked to bio-ontology terms in ways that enable users to verify the accuracy and relevance of data to the curators' interpretation. Curators hope in this way to acquire as much control as possible over the meaning attributed by users to each term and relation, and thus over the interpretation given to the claims about phenomena represented within bio-ontologies. Yet, what is interesting about bio-ontologies from the epistemic viewpoint is precisely the impossibility to exercise complete control over their interpretation and validity within an experimental context. Once the knowledge therein contained is brought back to the lab, the definitions provided by curators become one of several factors affecting users' interpretation of the data retrieved and of the claims about phenomena to which data have been connected. Unavoidably, experimentalists evaluate knowledge claims through the lens of their own expertise, their familiarity with instruments and materials, their specific background and interests, and of course their own findings. Far from being disruptive to curators' work, this fusion of users' local interpretation with the curators' efforts to globalise knowledge is what makes bio-ontologies into a unique and precious research tool. In their attempt to capture biological knowledge as it is produced in the lab, bio-ontologies act as a bridge between the context of experimental research, which produces and re-uses data towards discovery, and the context of bioinformatics, in which data are processed and classified so as to be disseminated as widely as possible. Such a bridge is indispensable within data-intensive research practices, where the iteration between data analysis and experimental research is as difficult to achieve as it is indispensable to scientific progress. Without curators' effort to globalise the knowledge expressed within bio-ontologies, computational tools such as databases would not succeed in making data accessible across research contexts and disciplines. Curators' judgment in selecting terms, relations, relevant data and meta-data is crucial, as testified by the increasing professionalization of the skills and training needed to make those choices (Leonelli 2010; Chow-White and Garcia-Sanchos 2011); curators are also responsible for updates to the system in view of new scientific developments (Leonelli et al 2011). At the same time, the categories through which data are ordered and retrieved in bio-ontologies need to be appropriated by each user, who has to integrate that knowledge with her unique background and interests in order to be able to re-use the data and to eventually revise that very knowledge. Hence, the meaning of knowledge statements within bio-ontologies is determined at least as much by users as it is by the definitions provided by curators.

3. Bio-Ontologies as Classificatory Theories

Much philosophical work on the characteristics and role of scientific theories has hitherto been grounded on the intuition that constructing a theory implies introducing a new language to talk about the natural world. This idea resonates with the Kuhnian understanding of incommensurability among paradigms, according to which paradigm shifts involve language shifts and even when key concepts stay the same, their meaning tends to differ (Kuhn 1962). More recently, this view has been explicitly defended by Lindley Darden (2006). In this section, I argue that bio-ontologies constitute a significant counter-example to this intuition: they constitute a form of scientific theorising that has the potential to affect the direction and practice of experimental biology in the long term, and yet they do not aim to introduce new language into biological discourse, but rather attempt to gather and express consensus on what constitutes established knowledge. In other words, bio-ontologies are an instance of how theory can emerge from the attempt to classify entities in the world, rather than from the attempt to explain phenomena. Recognising this role for bio-ontologies involves accepting that descriptive predicates can all have theoretical value, regardless of the generality and the novelty of the concepts that they contain – regardless, that is, of whether they can be clearly differentiated from empirical observations and whether they challenge existing knowledge claims. As I intend to show, whether descriptive predicates constitute a form of theory depends on the role that they play within the life-cycle of experimental research. Indeed, I shall argue that recognising this role for bio-ontologies sheds light on the epistemic features of data-intensive science, and particularly on interactions between theory-making and experimentation that characterise this mode of research: viewing bio-ontologies as theories illustrates how the theoretical claims underlying data classification can only be interpreted through consultation of the evidence supporting those claims, and thus through familiarity with the experimental practices through which that evidence has been produced.

As a starting point, I use Mary Hesse's network model of scientific theories (1980). In that view, theories are defined as networks of interrelated terms. The meaning of each term depends both on the definition given to the phenomenon to which the term applies and on its relation to other terms. Relations among terms are expressed by way of law-like statements (for instance, Newtonian mechanics includes the terms 'mass' and 'force', which are related by the law-like statement 'force is proportional to mass times acceleration'). Theories can thus be propositionally expressed by enunciating the series of law-like statements defining the relations among the terms used to refer to phenomena. Within this framework, Hesse argues that 'there is no distinction in kind between a theoretical and an observation language' (1980, 108). What Hesse means is that scientific language is replete with descriptive statements, whose truth depends on two fundamental factors: (1) their relation to other statements used in science to describe related phenomena (what Hesse calls 'the coherence of the network of theory'); and (2) their relation to various kinds of evidence ('empirical input'; Hesse 1980, 97). These descriptive statements may vary greatly with respect to the scope of phenomena to which they may be applied; the degree of generality of the concepts that they incorporate; and their degree of abstraction from specific empirical instances. Nevertheless, there is no obvious way to divide those statements in two distinct epistemic classes, one 'theoretical' and one 'observational'. The consequence of this realisation is to state that all descriptive statements have a theoretical value, and can be treated as theory depending on their context of use.

I find that this picture of how theory works in experimental science neatly fits the case of bio-ontologies. Bio-ontologies are a network of terms defined by a predicate (the definition of the term) as well as by a specified set of empirical evidence (the data classified under that term). The terms are related so as to form descriptive sentences applicable to a variety of empirical situations, such as 'a nucleus is part of a cell'. The truth of such statements depends on the truth of the definitions provided for each of the terms used, on the relation stipulated between each statement and the rest of the network of which it is part, and on its relation with available evidence as gauged by scientists. In other words, descriptive sentences captured in bio-ontologies have the same epistemic role as testable hypotheses: they are theoretical statements whose validity and meaning can only be interpreted and assessed with reference to empirical evidence and the conditions under which that evidence has been produced. Some of the biologists with whom I discussed this view initially resisted my characterisation of bio-ontologies as forms of theory: they understood biological theories as needing to express something that is not yet empirically established or that is very new, rather than what is already known. At the same time, they agreed that bio-ontologies are supposed to include 'all we know' about biology, and thus fit a view of theory as expressing existing knowledge about biological phenomena (including, but not limited to, newly available knowledge). The following passage, taken from an email sent to me by the director of a major center for the elaboration of bio-ontologies, illustrates the point:

> 'the relationships we capture tend to be the more basic and universally accepted ones, not those that are currently evolving as biological knowledge is extended. Perhaps that is why biologists do not think of database design as the highest level of our field, analogous to a mathematical model for the universe. I think ontology structure is closer to a true conceptual model, but even there the relationships captured are not on the cutting edge of biology but those that are already generally agreed on and not controversial' (27 August 2004).

Hesse argues that the choice, use and modification of each term and relation expressed in a network depends on the empirical study of the phenomena that those terms and relations are supposed to describe, and thus on the complex processes of experimentation, intervention and classification involved in empirical research (Hesse 1980, 84; 1974, 4-26). Thus, the theoretical representation of knowledge is instrumental to the goals and needs of empirical research. Similarly, bio-ontologies consist of a string of accurately defined, basic terms with clearly outlined reference to specific phenomena. As I have stressed, the main motivation behind the development of bio-ontologies is to facilitate the integration of data about all aspects of the biology of organisms. This means that the nature of the data available, as well as their format and the methods used by the laboratories that collected them, inform all decisions about which terms to use, how to define them and how to relate them to each other. In line with Hesse's arguments, terms are not defined on the basis of purely theoretical considerations, but they are formulated to fit the investigations actually carried out by empirical scientists. Terms referring to entities and processes that are not under investigation, and therefore do not have data associated to them, are not incorporated in bio-ontologies; similarly, definitions are often formulated through the observational language used by empirical researchers.

This became very clear to me when attending a Gene Ontology Content Meeting dedicated specifically to the terms *metabolism* and *pathogen*, two notions that are notoriously difficult to define and treated differently depending on the subdiscipline of interest (Gene Ontology 2004). Discussion among the immunologists, geneticists, developmental and molecular biologists attending the meeting started from a tentative definition of the terms and progressed by correcting that initial definition so that it would fit counter-examples. Most counter-examples were provided *in the form of actual observations* from the bench or the field. For instance, the proposal that pathogens be treated as an independent category from organelles, which was popular among immunologists, was dismissed by ecologists and physiologists on the basis of cases of endosymbiosis where pathogens turn out to be both symbionts and parasites of the same organism. In these cases, it was argued that pathogens cannot be treated as a separate, independent category from other microscopic components of the host's cell, since they also play a role towards the well-functioning of the cell as a whole. According to specialists in endosymbiosis and its role towards plant development, these pathogens should therefore figure as ‘part_of’ the cell, rather than as a separate entity with no relations to it. Certain kinds of bacteria (such as nitrogen-fixing bacteria) can have at the same time mutualistic and parasitic associations with the plants that host them (see Vernon and Paracer 2000). Thus, on the basis of observed cases, a whole theoretical category (the one of ‘pathogen’) was modified to fit a different context and definition.

This example shows how one of the keys to the successful functioning of bio-ontologies is precisely the avoidance of distinctions between theoretical and observational language and related concerns. Bio-ontologies are meant to represent knowledge about existing phenomena as currently investigated, rather than to make sense of them through detailed explanations or daring new hypotheses. For that purpose, it does not matter that bio-ontologies do not, and were never intended to, introduce new theoretical terms to explain biological phenomena. Not all scientific theories operate in that way,[5] nor are they all meant as sweeping, general perspectives unifying a whole discipline (as in the well-known cases of the central dogma within genetics or of evolutionary theory). Bio-ontologies are better understood as what I shall call *classificatory theories*. These are theories that emerge from a classificatory effort, in the sense that they aim to represent the body of knowledge available in a given field so as to enable the dissemination and retrieval of research materials within it; are subject to systematic scrutiny and interpretation on the basis of empirical evidence; affect the ways in which research in that field is discussed and conducted in the long term; and – most importantly if we are to regard them as theories – express the conceptual significance of the results gathered through empirical research.

[5] Interestingly, Darden quotes Hesse’s older work on analogies as supporting her intuition about the novelty of theoretical language (Darden 2006, 150). As I have shown here, Hesse actually changed her mind on this point.

As I highlighted above, bio-ontologies are developed through complex mediation between curators aiming to produce general definitions of entities and biologists interested in preserving the local definition used for those entities within their own research context. They are thus developed as cross-disciplinary communication tools, representations of knowledge that are general enough to fit most research contexts in biology, yet specific enough to capture local nuances and semantic diversity across epistemic cultures. What is remarkable about this achievement is its context-dependence. Bio-ontologies are supposed to capture domains of knowledge which are constantly evolving, both internally and in their interaction with each other. It is no wonder that curators struggle to keep up with biological developments (Parkhill et al 2010); and one crucial reason for curators to worry about lagging behind is the foundational role played by bio-ontologies in guiding the interpretation of data found online. Researchers who use bio-ontologies for data retrieval implicitly accept, even if they might not be aware of it, the representation of biological entities and processes contained within bio-ontologies at the moment in which they are consulted. This can have dramatic effects on how data are used in subsequent research (e.g. Leonelli and Ankeny 2012). And it is because of this crucial role in expressing available knowledge of biological phenomena, and thus guiding and structuring subsequent research, that bio-ontologies are best regarded as a form of theory.

One important objection might be raised at this point: why should we regard bio-ontologies as theories, rather than as yet another important component of scientific practice which affects, and yet does not constitute, theoretical knowledge? This objection seems especially strong in light of recent literature on the role played by elements other than theory, such as models, experiments and instruments, in shaping scientific research. However, what needs to be noted is that bio-ontologies are not simply influencing knowledge-making practices in biology in the way that a type of model or instrument would. Bio-ontologies certainly have the potential to structure and direct experimental inquiry in data-intensive science. But the reason for their epistemic power is that they actually express, rather than simply affecting, the knowledge obtained through scientific research. The knowledge that they express is best understood in relation to the collection of models, instruments and commitments made by the researchers who produced it, as in the case of any scientific theory;[6] and yet, that knowledge cannot be reduced to any of those elements of scientific practice, and actually provides a way to link and evaluate the epistemic results of using specific methods and tools to research nature. By virtue of their structure, their irreducibility to specific models (or even clusters of models) and, above all, their capacity to express scientific knowledge of biological phenomena in a way that summarises and directs empirical results, bio-ontologies are best viewed as theories rather than as scientific models or instruments.

This view of theory is certainly different from the sets of equations and axioms traditionally discussed in the context of the physical sciences. It is, however, not a newcomer within the philosophy of science. It parallels Claude Bernard's idea of theory as a 'stairway' towards the establishment of new scientific facts: 'by climbing, science widens its horizons more and more, because theories embody and necessarily include proportionately more facts as they advance' (1855, 165). Theoretical claims in this view are what brings us from one observation to the next, and make it possible to develop complex explanations.[7] Understood in this way, descriptive

[6] The importance of models in understanding theories has been widely discussed, for instance in recent work by James Griesemer (e.g. 2006) and by contributors to de Regt, Leonelli and Eigner (2009).

[7] In a similar vein, Kenneth Schaffner has offered a sophisticated view of 'middle-range theories' as devices for knowledge representation and discovery, in which he even briefly explored the usefulness of object-oriented approaches to programming (the ancestors of bio-ontologies) as a way of expressing theoretical commitments (1993). I see his approach as very congenial to mine, yet I am not considering it in detail here because of his different emphasis and targets (the role of law-like statements in biology and their relations to theories in physics and chemistry).

statements such as used to classify biological organisms within the taxonomic tradition could be seen as theoretical. This claim constitutes a step further than previous claims about the profound conceptual implications of specific taxonomies, especially concerning the conceptualization of organisms in biology, as put forward chiefly by John Dupré (1993, 2001 and, jointly with O'Malley, 2007). While Dupré argues that the classification of species and biological individuals has huge consequences for biological theory, I am arguing here that such classifications may sometimes even *constitute* theory in biology. It is important to note, however, that I am not arguing that all scientific classifications should be regarded as theories.[8] Rather, I am arguing that some classifications can play the role of theories depending on the extent to which they embody and express a specific interpretation of the overall significance of a set of empirical results – as in the case of bio-ontologies. Whether other classifications, such as the ones used in the natural history tradition or contemporary taxonomy, fit this requirement remains a topic for further debate. Staffan Mueller-Wille's reading of the classificatory practices used by Linnaeus, which he views as guided by the need to order and summarise data pouring in Linneaus' study from all corners of the globe, and yet resulting in a specific interpretation (an 'ontological scaffold') of the biological significance of those data, could indeed be interpreted as indicating that the Linnean classification system functioned as a theory in 18th century botany (Mueller-Wille 2007, with Charmantier 2012). Similarly, the arguments recently put forward by Bruno Strasser on the intertwining of experimental and classificatory practices in the construction of databases such as GenBank point to a potentially theoretical role of the classification of protein sequences in the context of 1970s-80s molecular biology (e.g. Strasser 2011).

4. Conclusions: Theories in Data-Driven Research

[8] I thank the editor for inviting me to stress this important point.

I have defended the view that bio-ontologies such as the Gene Ontology play the role of classificatory theories in data-intensive research modes. They consist of an explicitly formulated series of claims about biological phenomena, which is understood in relation to the methods, materials and instruments used to experiment on those phenomena and routinely adapted to scientific developments. On the one hand, they express the propositional knowledge that is relied upon when ordering data for the purpose of further dissemination and analysis, knowledge which expresses current understandings of the significance of data towards understanding biological phenomena and which can be challenged and modified depending on shifts in research contexts. On the other hand, their interpretation is grounded in tacit knowledge of the instruments, models and protocols characterising experimental research in biology. Bio-ontologies thus constitute an interesting example of how theory can emerge from classificatory practices in conjunction with experimental know-how. They express what is currently known about biological entities or processes, in order to further the study of those entities and processes through coordination among research projects and exchange of relevant data. They explicitly formulate knowledge that is taken to be widely assumed, yet is usually dispersed across publications and research groups. They need not be true or universal; rather, they capture the assumptions, interpretations and practices underlying the successful sharing and re-use of data within very specific contexts at a given moment in time.[9] This makes them into representations of the biological knowledge underlying data-intensive science.

This view has a number of far-reaching implications. One concerns the epistemic role of classification as a scientific method. At least since Francis Bacon's discussion of 'idols of the human mind', and passing through Pierre Duhem's work on theory as an aid to experimentation,

[9] This views contrasts with some aspects of the interpretation of ontology realism given by Smith and Ceuster (2010). A paper discussing the parallels and differences between these two views is in preparation.

classification systems have been recognised as unavoidably theory-laden. However, their role as active tools towards the exploration of the natural world and the discovery of new phenomena has been underestimated within the philosophy of science, despite the pioneering work by Ian Hacking (2002) and John Pickstone (2000) in highlighting the role of natural history as, respectively, a key style of reasoning and a way of doing in science. Consider for instance the debate on the existence and function of natural kinds, which can be read as focusing precisely on the implication of adopting specific classification systems to study the natural world. Traditionally, this debate has largely revolved around the issue of essentialism. Only recently has research focused on the ways in which scientists actually use natural kind terms in their research (Dupre 1993). Terms within bio-ontologies can be seen as 'kinds' in the sense suggested by Brigandt (2009) and Reydon (2010): that is, not as capturing some intrinsic and unchangeable feature of entities in the world, but rather as capturing scientists' current assumptions about what exists – the necessary (and yet revisable) background knowledge on which inquiry rests. Emphasising the epistemic role of bio-ontologies as theories involves highlighting the tight interrelation between the use and interpretation of specific classificatory categories, the set-up of experiments and the analysis of experimental results. Classification is, in this view, a key scientific activity that is actively pursued throughout the research process and has profound effects on its results and methods.

Another philosophical realm affected by my argument is, quite obviously, the conceptualisation of scientific theories. Over the last three decades, several philosophers have defended a deflationary view of theory broadly in line with Hesse's suggestions discussed above. Some of this work has focused on the nature of law-like statements (Cartwright 1983, Mitchell 2003) or on cases of scientific research where laws have arguably little or no role to play (Creager et al 2007). Some of the semantic conceptions of scientific theory, with their emphasis on various types of models (including material ones) collectively embodying theory, also fit this deflationary bandwagon (e.g. Giere 1999). Yet another relevant body of literature focuses on the nature of phenomena and their relationship to empirical reality (Bogen and Woodward 1988, McAllister 1997). The implications of my views for each of these debates need to be explored in detail. Claims within bio-ontologies can be interpreted as law-like in the phenomenological sense advocated by Cartwright (1983). And if, as argued by Bogen and Woodward (1988), phenomena are conceptualised as low-level descriptions of real entities and processes, they may amount to theories in the classificatory sense that I have defended, *depending on their function within any given research context*. Notably, semantic views of theories do not sit comfortably with this position, as argued for instance by Margaret Morrison (2000) and James Griesemer (e.g. 2006). Classificatory theory as embodied by bio-ontologies can be kept apart from the various types of models built to make biological sense of the data retrieved from databases (such as the visualisations of data proposed within model organism databases): bio-ontologies constitute the cluster of theoretical commitments on the basis of which objects and processes are tracked (to use Griesemer's terminology) and models of them can be built. My account is thus in line with the 'mediator' view of models (Morgan and Morrison 1999), according to which models are not themselves part of theories but rather function autonomously as bridges between theory (in my case, classificatory theory) and the world (biological entities and processes as targeted within bio-ontologies).

Thirdly, my views have implications for the debate on scientific realism. Prima facie, it might seem that viewing bio-ontologies as theories is in contrast with the realist commitments that bio-ontologies seem to express and that are often attributed to them by curators (for instance, when insisting that bio-ontologies represent 'the reality captured by the underlying biological science'; Hill et al 2008). I do not this that such conflict exists. Of course bio-ontologies do and should refer to real entities in the world. This does not mean that the knowledge we have of these entities is firm and infallible. Indeed, many scientists prefer to think of their knowledge as tentative rather than conclusive. A leading curator of a model organism database recently told me that she finds biologists to be put off by the assertive, unambiguous language used by bio-ontologies to express knowledge. Formulations such as ‘X is Y’ or ‘A regulates B’ are often interpreted by biologists as implying the unwarranted transformation of unverified statistical correlations into well-established causal claims. That absence of qualifiers and nuances makes biologists uncomfortable precisely because they know that several of these claims have not been conclusively proven and could very well turn out to be mistaken in the future. Thinking of bio-ontologies as theories takes nothing away from the curators’ realist commitments to providing the best possible representation of what is currently known in biology. Rather, this interpretation stresses that bio-ontologies are as fallible, dynamic and revisable expressions of biological knowledge, and that is indeed what makes it such an efficient tool for discovery in data-intensive science.[10] Viewing bio-ontologies as theories could actually help to ease the uneasiness felt by many biologists towards data-intensive methods, by uncovering the conceptual work underlying the development of databases and thus alerting biologists to the pitfalls of an uncritical use of these tools and to the need for constructive critique and expert feedback in order to improve their accuracy and reliability**.**

Finally, and coming back to the question initially raised in this paper, viewing bio-ontologies as a type of theory sheds light on the epistemic features of data-intensive science as a research mode. It illustrates how the procedures used to extract patterns from data are not purely inductive, but rather rely on a complex conceptual structure that is used to collect and retrieve data in the first place. It also shows how this process requires knowledge about the provenance of data and adequate tacit skills facilitating the use of that information to evaluate the claims fostered in bio-ontologies. Further, viewing bio-ontologies as theories facilitates an understanding of how crucial information technology is becoming not only to experimentation, but also to scientific reasoning. At the same time, emphasising the theoretical work embedded in computational data analysis constitutes a useful counter-point to the sometimes exaggerated role attributed to automation in data-intensive

[10] This view parallels the endorsement of scientific perspectivism in the analysis of data-intensive science presented by Callebaut (2012).

science (Evans and Rzhesky 2010). While digital infrastructure is crucial in shaping reasoning and 'random' searches, the interpretation of the biological significance of data is grounded in experimental knowledge and awareness of theoretical framing within bio-ontologies. Curators and experimentalists contribute interpretive input through manual interventions, and informal elements such as tacit skills and experimental familiarity with organisms play a key role in the assessment of the reliability of data and their evidential value. Database curators are the first to acknowledge that the re-use of data retrieved through automated searches stands in a complex relation to experimental and modelling practices. The automated analysis of datasets *in silico* provides the best opportunities for discovery when it is used in parallel to *in vivo* and *in vitro* experimentation.

Acknowledgments

This research was funded by the ESRC as part of the ESRC Centre for Genomics in Society; by the British Academy, as part of a Small Grant on 'Data-Driven Research in the Biomedical and Biological Sciences'; and by the ESRC/Leverhulme Trust grant for the project 'How Well Do "Facts" Travel?' (grant number F/07004/Z). For exceedingly helpful comments on previous drafts, I thank two anonymous referees, James McAllister, Jane Lomax, Maureen O'Malley, John Dupré, Thomas Reydon and Mary Morgan. Warm thanks for generative discussions are also due to Barry Barnes, Hans Radder, Giovanni Boniolo, Jim Griesemer and Werner Callebaut; the 'facts' group at the LSE; the participants in the workshops 'Making Small Facts Travel' (LSE, March 2008) and 'Data-driven research in the biological and biomedical sciences' (Exeter, April 2010); the audiences of seminars held at ISHPSSB 2007, Egenis, the HPS Department in Cambridge, the Konrad Lorenz Institute, EPSA 2009 and the SEMM; and Sue Rhee, who introduced me to this area of research.

References

ALLEN, J.F. (2001) Bioinformatics and Discovery: Induction Beckons Again, **Bioessays**, 23(1), pp. 104-107.

ASHBURNER, M. et al. (2000) Gene Ontology: Tool for the Unification of Biology, **Nature Reviews: Genetics**, 25, pp. 25-29.

BACLAWSKI, K. and NIU, T. (2006) **Ontologies for Bioinformatics** (Cambridge, MIT Press).

BARD, J.B.L. and RHEE, S.Y. (2004) Ontologies in Biology: Design, Applications and Future Challenges, **Nature Reviews: Genetics**, 5, pp. 213-222.

BERNARD, C. (1855/1927) **An Introduction to the Study of Experimental Medicine** (Mineola, Dover Publications).

BLAKE, J.A. and BULT, C.J. (2005) Beyond the Data Deluge: Data Integration and Bio-Ontologies, **Journal of Biomedical Informatics**, 39(3), pp. 314-320.

BOGEN, J. and WOODWARD, J. (1988) Saving the Phenomena, **The Philosophical Review**, 97(3), pp. 303-352.

BRAZMA, A., KRESTYANINOVA, M., and SARKANS, U. (2006) Standards for Systems Biology, **Nature Reviews: Genetics**, 7, pp. 593-605.

BRIGANDT, I. (2009) Natural Kinds in Evolution and Systematics: Metaphysical and Epistemological Considerations, **Acta Biotheoretica**, 57, pp. 77–97.

BROOKSBANK, C. and QUACKENBUSH, J. (2006) Data Standards: A Call to Action, **OMICS: A Journal of Integrative Biology**, 10(2), pp. 94-99.

BUETOW, K.H. (2005) Cyberinfrastructure: Empowering a 'Third Way' in Biomedical Research, **Science**, 308(5723), pp. 821-824.

BULT, C. J. (2006). From Information to Understanding: The Role of Model Organism Databases in Comparative and Functional Genomics. **Animal Genetics**, 37(Suppl 1), pp. 28–40.

CALLEBAUT, W. (2012) Scientific Perspectivism: A Philosopher of Science's Response to the Challenge of Big Data Biology, **Studies in the History and the Philosophy of the Biological and Biomedical Science: Part C**.

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

CHIPMAN, K.C. and SINGH, A.K. (2009) Predicting Genetic Interactions with Random Walks on Biological Networks, **BMC Bioinformatics**, 10, p. 17.

CHOW-WHITE and GARCIA-SANCHOS, M. (2011) Global Genome Databases Bidirectional Shaping and Spaces of Convergence: Interactions between Biology and Computing from the First DNA Sequencers to Global Genome Databases. **Science, Technology and Human Values**. Online First.

CREAGER, A.N.H., LUNBECK, E., and NORTON WISE, M. (2007) **Science without Laws: Model Systems, Cases, Exemplary Narratives** (Chapel Hill, Duke University Press).

DARDEN, L. (2006) Theory Construction in Genetics, in: L. DARDEN (Ed.) **Reasoning in Biological Discoveries** (Cambridge, Cambridge University Press).

DE REGT, H., LEONELLI, S. and EIGNER, K. (2009) **Scientific Understanding: Philosophical Perspectives** (Pittsburgh, Pittsburgh University Press).

DUPRÉ, J. (1993) **The Disorder of Things** (Cambridge, Harvard University Press).

DUPRÉ, J. (2001) In Defence of Classification, **Studies in the History and Philosophy of the Biological and Biomedical Sciences**, 32, pp. 203-219.

DUPRÉ, J. and O'MALLEY, M.A. (2007) Metagenomics and Biological Ontology, **Studies in the History and Philosophy of the Biological and Biomedical Sciences**, 38, pp. 834-846.

EGANA ARANGUANEN, M, BECHOFER, S., LORD, P., SATTLER, U. and STEVENS, R. (2007) Understanding and Using the Meaning of Statements in a Bio-Ontology: Recasting the Gene Ontology in OWL. **BMC Bioinformatics** 8 (57).

EVANS, J. and RZHESKY, A. (2010) Machine Science. **Science,** 329 (5990), pp. 399-400.

GENE ONTOLOGY (2004) **Minutes of the Gene Ontology Content Meeting**, Stanford University, 28-29 August. Available at: http://www.geneontology.org/GO.meetings.shtml

GIERE, R.N. (1999) **Science Without Laws** (Chicago, Chicago University Press).

GRIESEMER, J.R. (2006) Theoretical Integration, Cooperation, and Theories as Tracking Devices, **Biological Theory**, 1, pp. 4-7.

HACKING, I. (2002) **Historical Ontology** (Cambridge, MA, Harvard University Press).

HAMID TIRMIZI, S., AITKEN, S., MOREIRA, D. A., MUNGALL, C., SEQUEDA, J., SHAH, N.H. and MIRANKE, D. P. (2011) Mapping between the OBO and OWL Ontology Languages. **Journal of Biomedical Semantics**, 2 (Suppl 1), p. S3.

HEY, T., TANSLEY, S., and TOLLE, K. (2009) **The Fourth Paradigm. Data-Intensive Scientific Discovery** (Redmond, Microsoft Research).

HESSE, M. (1974) **The Structure of Scientific Inference** (London and Basingstoke, Macmillan).

HESSE, M. (1980) **Revolutions and Recontructions in the Philosophy of Science** (Brighton, The Harvester Press).

HILL, D.P., SMITH, B., MCANDREWS-HILL, M.S., and BLAKE, J.A. (2008) Gene Ontology Annotations: What They Mean and Where They Come From, **BMC Bioinformatics**, 9(Suppl 5), p. S2.

HOEHNDORF, R., OELLRICH A., DUMONTIER, M., KELSO, J., REBHOLZ_SCHUHMANN, D. and HERR, H. (2010) Relations as Patterns: Bridging the Gap between OBO and OWL. **BMC Bioinformatics**, 11(441).

HOWE, D. et al. (2008) Big Data: The Future of Biocuration, **Nature**, 455, pp. 47-50.

KEET, C. M. (2010) Dependencies between Ontology Design Parameters. **International Journal of Metadata, Semantics and Ontologies,** 5(4), pp. 265–284.

KELL, D.B., and OLIVER, S.G. (2004) Here Is the Evidence, Now What Is the Hypothesis? The Complementary Roles of Inductive and Hypothesis-Driven Science in the Post-Genomic Era, **BioEssays**, 26(1), pp. 99-105.

KUHN, T.S. (1962) **The Structure of Scientific Revolutions** (Chicago, Chicago University Press).

LEONELLI, S. (2009) On the Locality of Data and Claims About Phenomena, **Philosophy of Science**, 76(5), pp. 737-749.

LEONELLI, S. (2010) Documenting the Emergence of Bio-Ontologies: Or, Why Researching Bioinformatics Requires HPSSB, **History and Philosophy of the Life Sciences**, 32(1), pp. 105-126.

LEONELLI, S., DIEHL, A.D., CHRISTIE, K.R., HARRIS, M.A. and LOMAX, J. (2011) How the Gene Ontology Evolves, **BMC Bioinformatics,** 12(1).

LEONELLI, S. and ANKENY, R.A. (2012) Re-Thinking Organisms: The Impact of Databases on Model Organism Research, **Studies in the History and Philosophy of the Biological and Biomedical Sciences: Part C**.

LEONELLI, S. (2012) Making Sense of Data-Intensive Research in the Biological and Biomedical Sciences. **Studies in the History and Philosophy of the Biological and Biomedical Sciences: Part C**.

LEWIS, S.E. (2004) Gene Ontology: Looking Backwards and Forwards. **Genome Biology,** 6 (1), p. 103.

MCALLISTER, J.W. (1997) Phenomena and Patterns in Data Sets, **Erkentniss**, 47, pp. 217-228.

MITCHELL, S. (2003) **Biological Complexity and Integrative Pluralism** (Cambridge, Cambridge University Press).

MORGAN, M.S., and MORRISON, M. (1999) **Models as Mediators** (Cambridge, Cambridge University Press).

MORRISON, M. (2000) **Unifying Scientific Theories** (Cambridge, Cambridge University Press).

MUELLER-WILLE, S. (2007) Collection and Collation: Theory and Practice of Linnaean Botany. **Studies in the History and the Philosophy of the Biological and Biomedical Sciences,** 38, pp. 541-562.

MUELLER-WILLE, S. and CHARMANTIER, I. (2011) Natural History and Information Overload: The Case of Linnaeus. **Studies in the History and the Philosophy of the Biological and Biomedical Sciences.**

O'MALLEY, M.A., ELLIOTT, K.C., HAUFE, C. and BURIAN, R.M. (2009) Philosophies of Funding. **Cell,** 138, pp. 611-615.

PARKHILL, J., BIRNEY, E. and KERSEY, P. (2010) Genomic Information Infrastructure after the Deluge. **Genome Biology,** 11(7), pp. 402-404.

PICKSTONE, J. (2000) **Ways of Knowing: A New History of Science, Technology and Medicine** (Chicago, Chicago University Press).

RENEAR, A.H. and PALMER, C.L. (2009) Strategic Reading, Ontologies, and the Future of Scientific Publishing. **Science**, 325(5942), pp. 828-32.

REYDON, T.A.C. (2010) Natural Kind Theory as a Tool for Philosophy of Science, in: M. SUÁREZ, M. DORATO, and M. RÉDEI (Eds.) **EPSA - Epistemology and Methodology of Science: Launch of the European Philosophy of Science Association** (Dordrecht, Springer).

RHEE, S.Y. and CROSBY, B. (2005) Biological Databases for Plant Research. **Plant Physiology**, 138, pp. 1-3.

SHAFFNER, K. F. (1993) **Discovery and Explanation in Biology and Medicine** (Chicago: University of Chicago Press).

SMITH, B. et al. (2005) Relations in Biomedical Ontologies. **Genome Biology** 6, R46.

SMITH, B. et al. (2007) The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration. **Nature Biotechnology**, 25(11), pp. 1251-1255.

SMITH, B. and CEUSTERS, W. (2010) Ontological Realism: A Methodology for Coordinated Evolution of Scientific Ontologies. **Applied Ontology**, 5, pp. 79–108.

STRASSER, B. (2011) The Experimenter's Museum: GenBank, Natural History, and the Moral Economies of Biomedicine, 1979-1982. **Isis**, 102, 1, pp. 60-96.

VERNON, A. and PARACER, S. (2000) **Symbiosis. An Introduction to Biological Associations** (Oxford, Oxford University Press).

WYLIE, A (2002). **Thinking From Things** (University of California Press).