과학철학 26권 2호 (2023) pp. 31-58

Data, Representation, and Evidential Values in Biology†

Jinyeong Gim‡

Leonelli (2016) suggested a relational view of data against a representational view by emphasizing data-centric biology rather than the theory-centric tradition in the philosophy of science. This is because the first view allows for data journeys across laboratories using public database resources, whereas the second does not. This paper examines Leonelli's strategies to defend the relational view of data. Contrary to Leonelli's intention, it indicates that her strategies led to unnecessary misunderstandings of the relationships among data, representation, and evidential values. It will be argued that evidential values of data are inevitably based on a representational feature and that it is better to reconcile both views of data as complementary rather than contrasting.

Keywords: Data, Representation, Evidential Values, Data-Centric Biology, Leonelli

† The first manuscript of this paper was written and presented while participating in a reading seminar on Leonelli's Data-Centric Biology hosted by Professor Yeongseo Yeo. I thank Insok Go, Yeongseo Yeo, Bongguk Kim, and Kyoungjoon Oh for their helpful comments on this manuscript. I also thank anonymous referees for their beneficial comments.

‡ Lecturer, College of Liberal Studies, Seoul National University, jinyeong.gim@gmail.com

1. Introduction

What are data? What role do data play in biological sciences? These questions were discussed in Sabina Leonelli's book *Data-Centric Biology: A Philosophical Study* in 2016, which won the Lakatos Award. She (2016, pp.73-74) diagnoses general ideas about data under a theory-centric understanding of scientific knowledge in the philosophy of science, the so-called representational view of data, such that (i) data are representations of the world, (ii) data are primarily meant to test and validate theories, and scientific methods can be safely assumed to guarantee their reliability in this role. However, Leonelli noted that this view fails to capture the extensive efforts to disseminate and reuse data through Open Data Resources, Big Data infrastructure, data analysis via AI, and so forth in biology. She proposed that what counts as data depends on who uses it, how, and for which purposes (Leonelli 2016, p.78). She suggests a relational view of data under a data-centric understanding of biological knowledge by shrinking the discussion boundary from the overall sciences to biology. According to the relational view, (i) data are treated as potential evidence for one or more claims about phenomena, and (ii) data are formatted and handled in ways that enable its circulation among individuals or groups for analysis.

Interestingly, Leonelli argued that the representational view of data is *incompatible* with the relational view (Leonelli 2016, p.78). Instead of searching for the intrinsic features or functions of data, such as a local or idiosyncratic feature of data to particular experimental contexts, Leonelli insists that "judgments on which objects best work as evidence depends upon the preferences of the researchers in question, the nature of the claims under considerations, the materials (such as the organisms) with which they work, and the availability of other sources of evidence" (Leonelli 2016, p.78). Furthermore, Leonelli emphasizes portability of data as "a crucial precondition for using data as evidence" with her relational view of data (Leonelli 2016, p.80). For data to be traveled, first, data packaging procedures are essential to make data public resources, the so-called *decontextualization*. These procedures include data selection, formatting, standardization, classification, and the development of methods for retrieval, analysis, visualization, and quality control (Leonelli 2016, p.16). Additionally, for public data to be reused, *recontextualization* metadata is required. Curators generate metadata as a second type of label, providing information about the provenance of data (Leonelli 2016, p.29).

Leonelli's claims are philosophically striking because she focuses on new aspects of data that are often ignored in the philosophy of science. One critical aspect is the sociological context of data handling, such as the construction of databases by curators. Her new framework provides the possibility of understanding data independent of theoretical considerations. Nevertheless, all of her claims need to be more persuasive. Notably, although Leonelli argued for the incompatibility of a relational view of data with a representational view, both views rarely seem separate, unlike Leonelli's insistence. If data are not representative of the world, what would support the evidential role of data? The idea of data as potential evidence for something seems that both views are in concert.

I do not intend to defeat Leonelli's new interpretation of data. I agree that she skillfully guided us toward unexplored aspects of data. Notably, the mobilization or portability of data is a remarkable characteristic of data because it has not been sufficiently discussed in the twentieth century. Most philosophers of science consider data to be stable or objective grounds for evaluating theories. In contrast, Leonelli noticed a flexible or changeable characteristic of data depending on the research interests or contexts. To highlight this feature of data, she adopted strategies to contrast her relational view with the representational view. I admit that her strategies made us pay attention to her new ideas about data. However, they do not just attract our attention to data journeys across social contexts.

Unfortunately, her strategies seemed to lead to necessary questions about the relationships among data, representations, and evidential values of data. If both views of data are incompatible, can a relational view be established in the absence of a representational view? If data are not representations of the world, how do data become potential "evidence" for phenomena? Who advocates a representational view of data? Why do many philosophers of science regard data as neutral facts for testing or validating theories? Did they prohibit the circulation and multiple uses of data emphasized by Leonelli? I worry about misunderstanding data in the sciences because of Leonelli's strategies to defend her position by adding a proviso and the *incompatibility* of both views. I believe that she did not want to show the *wrongness* of the representational view. Instead, I believe that she makes people concentrate on overlooked aspects of data, such as the social procedures of data packaging, or reconsiders some unpromising data characteristics.

In this paper, I argue that Leonelli's view of data guides us toward a new direction or comprehensive framework and not as an alternative to the view that data are representations of the world. I indicate that both views of data must be reconciled to understand the nature and functions of data in biological sciences. It seems better to refer to the representational view of data as the *localized* view of data, as Leonelli's target. This is because she tackles the idea that data are fixed and mind-independent, not representations of the world. I am certain that Leonelli did not intend to reject representational data. Consequently, the incompatibility thesis should be eliminated or suppressed. The evidential value of data in science must rely on representational sources from the world before data are packaged and transported across research situations.

2. Leonelli's Strategies Against the Representational View of Data

(1) The Contrast Strategy

According to her analysis, many philosophers of science have dealt with data by testing scientific theories. For instance, Patrick Suppes (1962) discusses "models of data" in the epistemological context of comparing theoretical models and the world. Bas van Fraassen (2008, p.166) talks about a "data model" constructed from the analysis of the raw data. In this theory-centric context, Leonelli (2016, p.72) says there has been a tension of data between "viewing data as instances of the world and emphasizing their man-made nature." Most philosophers with a theory-centric stance regard data as neutral facts representing the world. However, theoretical ingredients are highly involved in empirical procedures such as observation, data production, and statistical interpretations. Notably, the theory of the ladenness of observation is a philosophical issue with the inductive method because of this tension. Assuming that a theory-centric tradition existed in the philosophy of science, Leonelli stipulated the traditional ideas of data as a representational view in the twentieth century. The main ideas of the representational view of data can be summarized as follows:

(1) **What are data?** Data are *representations* of the world.
(2) **What determines the content of data?** It is determined not by scientists, but by *nature*.
(3) **What features do data have?** (i) The content of data is *fixed*, regardless of how scientists use them. Therefore, (ii) data are *mind-independent*.
(4) **What role(s) do data play in sciences?** Data provide empirical grounds to *test* theories.

By contrast, Leonelli argues that data do not have fixed scientific value in and of themselves, nor can they be seen as mind-independent representations of a given phenomenon (Leonelli 2016, p.70). She defines data by *evidential value* ascribed to them at specific moments of inquiry (Leonelli 2016, p.70). Is this an evidential value? According to Leonelli, data are not limited to local laboratories but are disseminated through data packaging procedures in public institutions. Data are "any material product of research activities, which is treated as potential evidence for claims about phenomena and can be circulated across a community of individuals" (Leonelli 2016, p.195). Specifically, data are objects that satisfy the following two conditions (p.78): (i) data are treated as potential evidence for one or more claims about phenomena and (ii) data are formatted and handled in ways that enable their circulation among individuals or groups for analysis. Leonelli's relational view of data is as follows under the same questions above:

(1) **What are data?** Data are *potential evidence* for one or more claims about phenomena.
(2) **What determines the content of data?** This has been determined by both scientists and non-scientists.[1]
(3) **What features do data have?** (i) The data content is *circulated* depending on how scientists use it. Therefore, (ii) data are *mind-dependent*.
(4) **What role(s) do data play in sciences?** The roles of data are contextualized relative to the research *purposes*.

1) Leonelli (2016, fn.24 p.216) says about the content of data such that, "content is itself a function of the material features of data and of the expertise and skills of whoever interprets them." However, she does not pursue this in detail because "a thorough discussion of the relationship between data and information exceeds the scope of this book."

Leonelli sharply contrasts the relational and representational views of data. The two views answer the same question differently. Leonelli's position focuses on the data themselves rather than the epistemological context for the theory test. Her data-centric ideas are because there is no biological theory, such as Newton's laws, Einstein's principles of relativity, and Maxwell's equations. Leonelli mainly focused on sociological aspects, including the curator's activities in constructing databases and packaging data, which seem to rarely depend on the goal of theory choice. I admit that Leonelli successfully focused on newly unexplored topics in biology by adopting a contrasting strategy.

(2) The Incompatibility Strategy

Despite Leonelli's practical setting, we should be cautious regarding whether we must defeat the representational view of data by considering Leonelli's relational view. When suggesting a new data framework, which features does she want to emphasize? Does she deny all the answers from the representational view of data to the four questions above? Or does she partially answer and simultaneously reject others? This is because the differences between the two positions do not imply counter-examples for a specific position. In the face of this contrast, based on the four questions, both views of data are distinguished from each other. Further, Leonelli seems to endeavor to strongly emphasize the relational view of data by adding a provisional thesis to those contrasts: incompatibility.

> Many participants think that debates concerning data-intensive science should be grounded on a context-independent definition of what data are. This arches back to a representational view of data as entities that depict a specific part of reality independently of the circumstances under which they are considered. Under this interpretation, analyzing

> data involves uncovering which aspects of reality they document, and their epistemic significance stems from their ability to represent such aspects of reality irrespectively of the interests and situations of the people handling them. This view is *incompatible* with the idea that the same set of data can act as evidence for a variety of knowledge claims, depending on how they are interpreted— a feature that I take to be central to understanding the epistemic power of data as research components (Leonelli 2016, p.79, *an emphasis added*).

Leonelli sharply distinguishes her relational view from the representational view of data, asserting that these views cannot be in alliance. As the gap between the two views of data is more profound, this incompatibility thesis helps accentuate her position on the representational view. I think that Leonelli chose the notion of incompatibility for this purpose. However, the literal meaning of incompatibility indicates a contradictory relationship between the two views.

Irrespective of her original intention, this thesis raises fundamental questions regarding how the evidential values of data can be acquired based on the literal interpretation of Leonelli's claim. Is it possible to establish a relational view of data in the absence of a representational view as a whole? If data are not representations of the world, how do data become potential "evidence" for phenomena? Who advocates a representational view of data? Why do many philosophers of science regard data as neutral facts for testing or validating theories? Did they prohibit the circulation and multiple uses of data emphasized by Leonelli? These questions are related to the nature and function of data in science. Regardless of favoring a particular view, it is necessary to examine whether Leonelli's argument is adequate. This examination is essential for evaluating Leonelli's insights into data.[2)]

2) Of course, someone may think that the term incompatibility does not need to be considered seriously. This term might be used by Leonelli to show

In summary, Leonelli suggests a relational framework against a representational view of data. According to her main argument in the book, the representational view misleads understanding data in biology. In contrast, her relational view adequately guided us. Furthermore, the two views of data are mutually incompatible. Hence, she defends the relational view of data. I agree that Leonelli guides us toward new features of data in biology. However, should we accept her argument? In the next section, I scrutinize Leonelli's strategies for defending her position. She introduced two strategies: (i) contrasting different items between both views and (ii) exaggerating the incompatibility between both views. I show that we need not choose only one of the two views by suggesting a weakly modified interpretation of data in biological sciences.

3. Examining Leonelli's Strategies

(1) Differences Are Trivial

Who understands data based on a representational view? If Leonelli points out the philosophical flaws in the representational view of data, we can assume that someone defends this view. Interestingly, Leonelli did not specify the names that advocate this view. Instead, she briefly overviews the history of the philosophy of science in the twentieth century, focusing

---

where data's evidential values come from. Furthermore, I think that the incompatibility thesis must not be interpreted literally. This paper aims to prevent us from misunderstanding relations among data, representation, and evidential values. Particularly, I will argue that data's evidential values rely upon representational successes no matter how they are relative to research interests or contexts. Thus, I focused on Leonelli's strategies, not her philosophical claims.

on theory-centric issues. She (2016, p.74) insisted that a representational view of data is "tied to a theory-centric understanding of scientific knowledge production." The association between a representational view of data and a theory-centric understanding of scientific knowledge defines the function or role of data in science. When equipped with reliable data, we can test theoretical hypotheses. The more precise the data, the more reliable the data for knowledge claims. An inductive method guarantees that the more examples a hypothesis has, the more credible it is. Deductive methods guarantee that a theoretical hypothesis entails empirical data. In other words, data are resources that support observational statements, particularly in testing theoretical hypotheses.

There is no controversy over Leonelli's short overview of data in an old-fashioned philosophical project called logical empiricism. As Leonelli says (2016, p.71), many logical empiricists regard the term *data* as "what is given" etymologically. In addition, they may accept two features of data: (i) fixation and (ii) mind independence. Typical examples include Rudolph Carnap's confirmation theory and Karl Popper's falsification theory. These scientific methods make a vulnerable assumption: the distinction between theoretical and observational vocabulary (and statements). As Leonelli (2016, p.73) mentions, the theory-ladenness of observation aroused doubt about this distinction. Leonelli (2016, pp.74-75) also says that "this theory-centric view of scientific knowledge has been challenged over the past three decades by what is sometimes dubbed the *practice turn* within Anglo-American philosophy of science."

Why did logical empiricists seem to consider data fixed and mind-independent? As Leonelli points out, they concentrate exceptionally on epistemological issues. For example, Reichenbach insists that science philosophers must rationally understand how theories are empirically tested. It is well-known that he distinguishes the context of justification from that of discovery. The former context is a theory-centric stance on the sciences, and the nature or functions of data themselves were not centered on philosophical issues at that time. Data must be neutral facts within logical empiricism because they should not be changeable, relying on subjective intentions or expectations. The data content must be determined naturally through objective experiments or measurements. For this reason, data represent the world and not the subjective judgments of individual researchers.

Recall that Leonelli emphasized the portability of data rather than fixed immobility. She liberally allows human data manipulation when data are circulated from one laboratory to a public resource and vice versa. The content of data is flexible and depends on how researchers employ them. Therefore, according to the relational view, data depend on the research purpose. Here one question arises. Did logical empiricists normatively forbid data circulation, manufacturing, or packaging? Of course, no logical empiricist indicated their position on this question. This is because they are scarcely interested in Leonelli's concerns about data, and not because data represent the world. No matter how logical empiricists did not concentrate on possible aspects of data except the theory-test context, and no matter how they presumed that data content is determined not by the subjective mind but by an objective nature, their ignorance of Leonelli's interests is immune to criticism. No logical empiricist denied the fixation and mind-independence of data. The social aspect of data just was out of issues in the logical empiricism.

Notice that Leonelli never denies that data are representations of the world, which is the main idea behind the representational view of data. She rejects subsidiary data features from the representational view: (i) fixation or idiosyncrasy to experiments, and (ii) mind-independence. Probably, logical empiricists regard data as fixed and mind-independent to obtain objectivity. They did not focus on social features of data, such as circulation or portability, because they concentrated on the epistemological context, including confirmation or falsification. They implicitly assumed that the objectivity of data could be obtained from common results through diverse experiments on the same objects worldwide, independently of subjectively biased judgments. For this reason, it is better to refer to Leonelli's target not as the representational view of data as a whole but as a *localized* view because she emphasizes a new viewpoint of data beyond the idea that data are locally idiosyncratic to specific experimental contexts. The main target of the Leonelli attacks is the intrinsic locality of data and not the representational feature. As discussed later, Leonelli assumed that data represent natural phenomena in her book. That is, it will reveal that the representational features of data are immune to Leonelli's criticism. Nonetheless, because of Leonelli's incompatibility thesis, we must choose one of the two views of data. However, suppose Leonelli's original intention in contrasting the two views was to urge attention to new issues of data in a theory-centric stance. In this case, we do not need to consider the incompatibility of the two views. However, this was not a matter of choice. The theory-test context is just one of the diverse uses of data. Leonelli aided in shifting attention from an epistemological context of data.

(2) Incompatibility Is Unnecessary

Here, I discuss this incompatibility thesis. The term *incompatibility* between the two views implies that one view is correct, whereas the other is wrong.[3)] If we interpret this thesis literally, we are forced to choose only one position. I think that this interpretation is not Leonelli's original

3) In this discussion on 'incompatibility,' I assume there is no case such that two views simultaneously are false. I assume the following cases, (i) both views are correct, (ii) the representational view is incorrect, whereas the relational view is correct, and vice versa. Thank a reviewer's helpful advice.

intention. I wish to interpret this thesis as a tool for Leonelli to emphasize her new perspective on data in the biological sciences. I emphasize this weak interpretation of the incompatibility thesis. Simultaneously, I take precautions against a solid or literal interpretation of this thesis, such that both views of data must not be in alliance with each other.

In the previous section, I suggested that the two contrasts between the two views of data, (i) idiosyncrasy or locality of data to experimental environments, and (ii) mind-independence of data, are trivial features of data in the representational view. The sharp distinction between these two views concerns the nature of data. In the representational view, data are presented as representations of the world. According to the relational view, data provide potential evidence for diverse claims about the phenomena. However, the question is whether these definitions are competitive or contradictory. Based on a literal interpretation of the incompatibility thesis, the answer to this question may be yes. However, we did not address this issue.

Please remember that Leonelli described the representational view of data in two ways: (i) the nature of data and (ii) the role of data. Should the representational view of data be tied to a theory-centric understanding of scientific knowledge? Can the representational view remain independent of a theory-centric understanding? Can the two aspects of the representational view of data be separated? Leonelli seems to believe that these two aspects are inseparable. Logical empiricists were just interested in data in the context of the theory test. To understand the nature and role of data, we first need to consider James Woodward's view of data, instead of logical empiricism. Woodward was a philosopher of science who philosophically guided us on the importance of data. A crack between the two aspects appeared in the late 1980s when Bogen and Woodward (1988) published *Saving the Phenomena.*

Woodward and Bogen focused on distinguishing data from phenomena (see also Woodward 1989). They stressed that data should be distinguished from these phenomena. Phenomena are "relatively stable and general features of the world which are potential objects of explanation and prediction by general theory" (Woodward 1989, p.393). In contrast, data are "what registers on a measurement or recording device in a form accessible to the human perceptual system, and to public inspection" (Woodward 1989, pp.393-394). Compared with phenomena, data are "not viewed as potential objects of explanation by or derivation from general theory" but regarded as "evidence for the existence of phenomena" (Woodward 1989, p.394). That is, data are representations of phenomena because "phenomena are detected through the use of data" (Bogen and Woodward 1988, p.306). This distinction was the beginning of the understanding of the nature and functional roles of data *in themselves*. It should be noted that the evidential role of data concerns claims for phenomena, not theoretical explanations or predictions.

Bogen and Woodward argued that theories are directly related to phenomena rather than to data. Logical empiricists and Popper concentrated on the relationship between theory and observational data. Advocates of an inductive method suggest that inductive inference is always an inference to the best explanation because believing a theoretical claim, *T*, based on evidence *E* must always take the form of a demonstration that *T* figures in the best explanation of *E* (Woodward 1989, p.398). In contrast, advocates of a deductive method, such as Karl Popper, or a deductive-nomological model of scientific explanation, such as Carl Hempel, suggest either that claim *T* is falsified by evidence that claims *T* entails evidence *E*, or that claim *T* entails evidence *E* (see Leonell 2016, p.74). Both logical relationships between theories and data suggest that data are directly related to theories, either methodologically or explanatory. Woodward provided a method to consider data in the absence of a connection with theories by distinguishing data from phenomena.

Woodward renounced the functional role of data as evidence for testing theoretical hypotheses. According to Woodward, data play an evidential role in demonstrating the existence of phenomena that are explained or predicted by theories rather than data.

> One of my central claims is that one can justifiably believe that data provide reliable evidence for some phenomenon *without* being in a position to explain or derive facts about the data and *without* understanding in detail the causal mechanisms by which the data are produced (Woodward 1989, p.398).

That is, data are evidence not for theoretical hypotheses, but for phenomena. He, along with Bogen, emphasized the distinction between phenomena and data to criticize logical empiricists' view that data are empirical evidence to test or validate theoretical hypotheses. Woodward argues that theories are related to unobservable phenomena explanatorily and that data indicate the existence of the phenomena.[4] Woodward stressed the importance of data independent of the context of the theoretical choice. He regarded data as representations of the existence of phenomena. This is the main idea behind the representational view of data. Interestingly, Leonelli confesses that "Bogen and Woodward's work constitutes a crucial reference point for my account" (2016, p.84). Leonelli states that Bogen and Woodward's analysis shares her interest in using scientists' concerns and actions as the beginning of understanding the functional role of data in providing evidence of claims. Leonelli (2016, p.88) also says

4) Woodward says: "What matters in connection with the relationship between data and phenomena is not that one be able to produce derivations or detailed causal explanations of the data but that the data should be *reliable evidence* for the phenomena in question" (Woodward 1989, p.398, emphases original).

this explicitly: "The emphasis in this [representational] account is on the crucial role of statistics in helping scientists to fit scattered data points into a significant pattern, which can be used to test theoretical predictions or, in Bogen and Woodward's terms, to corroborate claims about phenomena."

Thus, I have the question: Is it possible for data to become potential evidence without regard for not being the representational role of data? Certainly, impossible! Science is the study of the world. Theoretical predictions must be compared with those expected to represent the target system. In the absence of theories, in the case of molecular biology, biologists infer a regular pattern to represent the structural features of a molecule based on the distribution of spots from X-ray diffraction technology. The discovery of DNA structures by Watson and Crick is a typical example. Theoretical and instrumental inquiries in science essentially require a representation of the world. These data form an empirical foundation for investigating the world. All evidential values in science inevitably originate from the representational feature of data. Leonelli (2016, p.85) also accepts that "data carry information about what the world is like." We can presume that Leonelli never persistently refuses the idea that data are representative of the world. Consequently, Leonelli did not aim to defeat all the aspects of the representational view of data. She questioned some aspects of this view.[5)]

Leonelli's definition of the representational view of data stems from an outdated philosophical stance, logical empiricism. If we follow Leonelli's definition of the representational view, her alternative, relational view, seems more persuasive because logical empiricism is obsolete. However, data need not be considered in a theory-centric understanding of scientific

5) A reviewer points out that I need to address the problem that arises when visualized data in science does not account for how it represents the world. I appreciate the reviewer's comment to clarify my claim.

knowledge production with a logical empiricist's viewpoint. Woodward, along with Bogen, provided a critical cornerstone for counting data as evidence for the existence of phenomena without regard to either testing theories or explaining (or predicting) phenomena. Consequently, when adopting Woodward's perspective, the representational view of data is no longer necessarily associated with a theory-centric understanding of knowledge that depends on scientific methods. Leonelli's relational view of data implicitly assumes that data represent the world. She depends heavily on Bogen and Woodward's position on the distinction between data and phenomena. Without the representational features of data, they play no role in supporting the hypotheses. Therefore, the incompatibility thesis must not be exaggerated.

4. Reconciling the Representational with Relational Views of Data

Until now, data have been representative and potentially evidential for various claims regarding these phenomena. This is the ideal answer to the following question: What are data? Thus, the last question remains: How can these views be reconciled? Recall that Leonelli's contrasts between the two views of data relate to the following questions: (i) What determines the content of data? (ii) What features do data have? (iii) What roles do data play in the biological sciences? Let us answer those questions in turn.

Related to the first question, Leonelli argues that both views are crucially different. According to the representational view, data content is determined only by nature. However, Leonelli stressed that it must be determined by nature and human agency, including database curators. Leonelli emphasized that data can be formatted diversely. Leonelli provided the following three examples. Figure 1 shows the positions of the gene markers on the chromosome. The region on the third chromosome of *Arabidopsis thaliana* includes indications for known genes and the proteins they encode, which are stored in the TAIR database and viewed through GBrowse. Figure 2 shows the scattered colors indicating the gene expression levels in a microarray cluster. The mouse cDNA microarray contained approximately 8,700 gene sequences and an interpretation chart of the relative expression levels. Figure 3 shows photographs captured to document different stages of embryological development. Charts of base pairings, two-dimensional distributions of genes, and visualizations of developmental gene portions show different types of data depending on different interests or research goals.

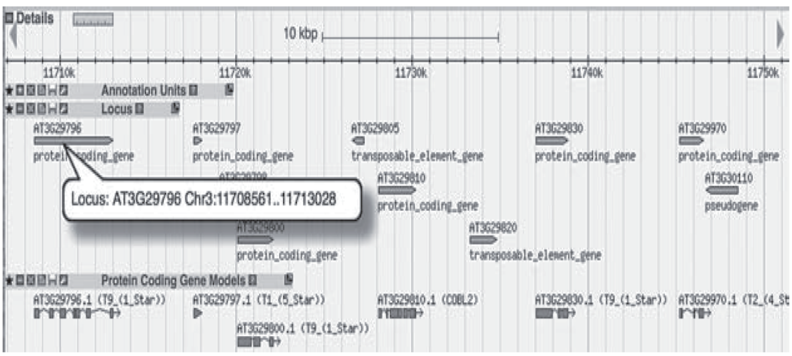

Figure 1: A 45kbp-wide region (Leonell 2016, p.71)

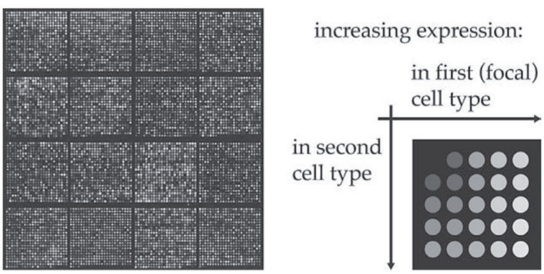

Figure 2: A mouse cDNA microarray (Leonell 2016, p.72)

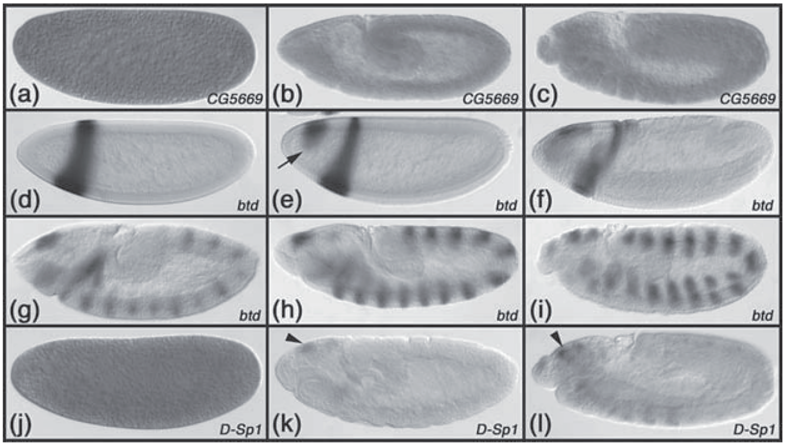

Figure 3: Photographs of fruit fly embryos (Leonell 2016, p.73)

Be cautious about how much data content is determined by human agency. Although the three examples are different, the shared content of the three figures represents genetic expression. I suggest a distinction in the content of data between structural properties and visualized forms. Data on the structural properties of phenomena include concrete sequences of bases in genomes, the number of base pairings, and relative locations of genetic expressions. These contents are determined by nature and are independent of the scientists' interests or research goals. Scientists' perspectives are merely engaged in the issue of how to visualize the structural properties of phenomena. Figure 1 shows the mapping relationships between the genes and proteins. Figure 2 shows the colorful differences between the normal and abnormal genetic portions. Figure 3 shows the relative locations of the Sp genes across sequential developmental stages. The visualized forms determine the different physical properties. Figure 1 shows the number of base pairs. Figure 2 shows the distribution of these genes. Figure 3 shows the temporal stages of the development. The structural properties of phenomena are independent of scientists' interests, whereas the visualized physical properties hinge on human agency intentions. Not all data content is solely determined by nature or scientists. Both nature and scientists are engaged with various types of data.

Regardless of what determines the content of data, it can be circulated through diverse forms, depending on the researchers' goals. It is important to note that the representational features of data regarding the structural properties of phenomena are still mind-independent. This is because these properties are naturally determined. Scientists' interests also influence various physical features of data. Different features of data are based on different data content. Consequently, data can be reused and circulated through diverse research goals. Simultaneously, data on the structural properties are invariant and mind-independent.

Finally, we consider the role(s) of data. Recall that both views acknowledge the evidential role of data. The main difference between the two views is whether data travel across laboratories. Leonelli referred to this gap as *locality*. She argued that the representational view regards data as idiosyncratic to experimental contexts, whereas the relational view does not. Leonelli implies that the evidential value of data in biology has yet to be previously determined. Essentially, Leonelli criticizes a representational view of data because this view neglects the portability of data as potential evidence across laboratories by saying that "for Bogen and Woodward, data are local evidence for nonlocal claims" (Leonelli 2016, p.86).

When Leonelli advocates the term *portability* of data under her alternative, a relational view of data, to a representational view, she points out that data as representations of the world are locally situated in the specific experimental circumstances that produce the data. It seems that Leonelli overlooked the distinction between data *dependence* on the experimental apparatus and data *immobility*. In the former, the data content is counted as a type of procedure for producing the data. If several instruments are used to manipulate the same object, the data on the target object are proportional to the number of instruments. Woodward stated that phenomena are relatively stable, whereas data are idiosyncratic, showing the dependence of data on ways to approach the world. However, the latter implies that if a datum is used as evidence within a specific context, it cannot be used as evidence within another.

The dependence of data on material circumstances is not identical to the immobility of data. Regardless of how data are produced in a specific area for a particular purpose, dependence never prohibits data from traveling or traveling from one laboratory to another. Furthermore, Woodward does not intend to emphasize the immobility of data by describing locality.

> Data are idiosyncratic to particular experimental contexts and typically cannot occur outside of those contexts. (...) Phenomena, by contrast, are not idiosyncratic to specific experimental contexts. We expect phenomena to have stable, repeatable characteristics that will be detectable using various procedures, which may yield quite different kinds of data (Bogen and Woodward 1988, p.317).

This indicates that diverse methods can represent a single phenomenon. That is, the more significant the type of instrument, the more data. Woodward's locality does not imply data immobility.

Recall that Leonelli insisted on data-centric biology through a relational view of data. Her position is the counterpart of a theory-centric understanding of scientific knowledge in conjunction with a representational view of data. However, this comparison requires further investigation. According to Woodward, data are evidence of phenomena, but they can also be handled far from a theory-centric approach. In addition, if the locality of data does not entail immobility from Woodward's view, Leonelli's relational view of data no longer serves as an alternative to the representational view. Of course, Leonelli's analysis of data packaging demonstrates a new kind of data journey through public databases to which philosophers of science have recently paid attention. Nonetheless, her new analysis of the socioeconomic aspects of data dissemination needs to provide more adequate arguments to refute the representational view of data. The evidential values of data are definitely based on the structural and visualized properties of the phenomena. Consequently, the previous four questions about data can be answered again when reconciling both views of data as follows:

(1) **What are data?** Data are *representations* of the world. For this reason, data provide *potential evidence* for another claim about phenomena.
(2) **What determines the content of data?** There were two types of data content: (i) The content of data on *structural properties* of the phenomena is determined by nature. (ii) Researchers' interests determine the content of data on *physical forms* of the phenomena.
(3) **What features do data have?** Data content *may* be circulated in diverse formats, relative to mind-dependent usage.
(4) **What role(s) do data play in sciences?** One of the main roles of these data is to provide empirical grounds for testing theories in diverse functional contexts.

In short, I focused on the two views of data by examining the nature, content, and functional roles of data. I stress the epistemological ground of data representing the world before data travel. I think that the evidential values of data are never solely determined by experimental circumstances or the researchers' aims in investigations. Certainly, data are initially given by *the world*. The representational nature of data is the fundamental basis for any evidential claim in science. I hope for the compatibility of both views of data because the relational view fundamentally presumes that data in science are representations of the world. Depending on the research context, different evidential values for the same data are possible. However, the only reason that data can be evidence is that they represent the world. Hence, Leonelli's relational view must be based on the representational view.

Some may complain that my claim is never new because Leonelli has already discussed two practices of science.[6)] Is Leonelli arguing about the compatibility of these two views? In my opinion, she does not insist on any compatible unification between the two different views of data but instead emphasizes data packaging rather than modeling in Subsection 3.4. By citing Paul Edwards' and Patrick Suppes' discussions, she highlights data handling practices, such as data formatting and the choice of metadata (See Leonelli 2016, pp.88-90). The only one with which data and models share is that both are manipulated "in their quest for knowledge" (Leonelli

6) A reviewer remarked, "Leonelli clarifies the difference between data packaging and scientific modeling in subsection 3.4 of her book. She (2016, p. 91) argues that both data and models are objects researchers continuously manipulate and intervene on in their quest for knowledge. But modeling and data packaging are different processes using different scientific tools. Leonelli's data and scientific models as structural representations are compatible but distinct scientific practices. If the author's argument concerns the compatibility between Leonelli's data and models, then this argument simply repeats what has already been fully explained by Leonelli (2016)." I appreciate this comment in that it is helpful to reconsider the relationship between Leonelli's relational view of data and scientific practice. Clearly, I do not focus on different practices, modeling of data vs. data journeys, but different views of data, representational vs. relational. My interest is not *practical activities* in which data are used but the (functional) *nature* of data. Of course, the latter is related to the former, but both are different issues. As I mentioned in the Introduction, Leonelli opens an unexplored social context of data. Independent of contexts of practices, I concentrate on whether or not evidential values of data can be possible without the representational contents of data. My answer is negative!

2016, p.91). Scientific data and models aid in the acquisition of scientific knowledge. I think that it is nonsense to judge that Leonelli pursued the compatibility of the two views of data based on this general idea.

In contrast, Leonelli attempts to differentiate her concern with data from either theory-centric research or representational visualizations through modeling. When reviewing Edwards' discussion, she (2016, p.90) says that in "both in the climate and life sciences, the packing of data is *not solely or even always* focused on statistical analysis; it *also* involves work on the medium, format, and order of data, as well as the ways in which they are combined with other datasets or selected/eliminated in order to fit a certain kind of vehicle (*emphases added*)." The statistical analysis indicated Suppes' remarks on data models in an epistemological context. She (2016, p.91) repeatedly argues that "the difference between data and models lies in the goals ascribed to those manipulations and the constraints under which they take place. The same object can function as data or as a model, depending on whether it is intended to function as evidence for claims or to teach researchers something about the world." Leonelli endeavored to distinguish evidential functions from representations of the world. That is, packaging and modeling both require the intervention and manipulation of data but for *different* purposes. Modeling is performed with the explicit goal of revealing specific characteristics of the world, and representational results serve as a bridge between the world and scientific knowledge. On the other hand, packaging aims to increase data portability, making it more likely to be used as evidence. According to Leonelli (2016, p.90), interpretive constraints in modeling are not goals when constructing metadata and data formatting but somewhat additional effects in securing mobility. It seems that Leonelli separates modeling practices, such as building models of data, from a packaging practice for data journeys, as she emphasizes the difference between the representational and relational views of data. Leonelli consistently employs her strategy to differentiate a context of representing the world from that of socially circulated data. But I do not argue for the compatibility of the different *practices*. I argue that evidential values of data depending on the researcher's interest, cannot be established in science without representational data.[7)]

5. Conclusion

Leonelli defended her relational view of data from a representational perspective. According to Leonelli, the representational view of data deals with data having intrinsic features, such that they are situated, local, and idiosyncratic to specific experimental contexts. However, the evidential value of data must be determined and depends on how researchers use data for their purposes. Therefore, Leonelli advocates a relational view of data and argues that both views are incompatible.

However, Leonelli's definition of the representational view of data must be revised. Leonelli compared both views on whether data are local or mind-dependent. Woodward first mentioned the local aspect of data, not logical empiricism. Woodward dealt with data independent of theory. Woodward adhered to a representational view of data. Interestingly, Leonelli accepted Woodward's fundamental distinction between data

7) In this paper, the term 'compatibility' has been used in the sense that Leonelli's relational view of data can never be established by excluding the representational view. This compatibility is asserted because it acknowledges that the context in which data are used in science is not only modeling to represent the world but also packaging in which data are used for many purposes. As I repeatedly acknowledge Leonelli's efforts to launch unexplored issues of data, from the etymological viewpoint of data, which means 'data is given,' I allow for the significance of representational data and relational contexts.

and phenomena. Additionally, the mind-independence of data is not a central feature of the representational view because no one strongly argues for it. The core of the representational view is that data are the outcomes of scientific investigations that represent the world. Therefore, The Woodwardian representational view of data is no longer associated with a theory-centric understanding of scientific knowledge. If we adopt Woodward's view of data, the representational view is the foundation of Leonelli's view of data. It is reasonable to think of Leonelli's strategies for not defeating the idea of data as representations of the world, but paying attention to new issues of data in biological sciences.

Conclusively, let me respond to the following initial questions: What is the nature of scientific data, and how do data function in scientific practice? According to the Woodwardian version of the representational view of data, data represent phenomena. Data play an important role in identifying phenomena. The evidential value of data did not depend on the local laboratory environment. The content of data on the structural properties of phenomena is determined by nature. The content of data on the physical forms of phenomena is influenced by human agency. The Woodwardian version of the representational view of data can be reconciled with Leonelli's relational view because the locally idiosyncratic feature of data is not a necessary condition of the representational view.

References

Bogen, J. and Woodward, J. (1988), "Saving the Phenomena", *The Philosophical Review*, 97(3): pp. 303–52.

Leonelli, S. (2016), *Data-Centric Biology: A Philosophical Study*, The University of Chicago Press.

Suppes, P. (1962), "Models of Data", in Ernest Nagel, A. T. and Suppes, P. (eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, Stanford, CA: Stanford University Press, pp. 252–61.

Van Fraassen, B. C. (2008), *Scientific Representation: Paradoxes of Perspective*, Oxford University Press.

Woodward, J. (1989), "Data and Phenomena", *Synthese* 79(3): pp. 393–472.

논문 투고일	2023. 06. 20.
심사 완료일	2023. 07. 11.
게재 확정일	2023. 07. 26.

생물학에서의 데이터, 표상, 그리고 증거적 가치

김 진 영

레오넬리(2016)는 과학철학에서의 이론 중심 전통보다 데이터 중심의 생물학을 강조함으로써 데이터의 표상적 견해에 거부하고 관계적 견해를 제안했다. 왜냐하면 관계적 견해는 공용 데이터베이스를 사용하여 실험실 간 데이터 이동을 허용하지만 표상적 견해는 그렇지 않기 때문이다. 그러나 데이터의 표상적 견해가 과연 무엇인지, 그리고 데이터의 본성에 대한 이해를 추구할 때 우리가 표상적 견해와 관계적 견해 중에서 택일해야 하는 것처럼 두 견해를 대조시키는 레오넬리의 전략이 적절한지 검토될 필요가 있다. 레오넬리의 의도와 달리, 그녀의 전략은 데이터, 표상 및 데이터의 증거적 가치 사이의 관계에 대한 불필요한 오해를 초래할 우려가 있다. 데이터의 증거적 가치는 필연적으로 표상적 특징에 기반해야 하며, 데이터에 대한 두 가지 견해를 대조하는 것보다 상호 보완적인 것으로 간주하는 것이 더 적절하다고 주장할 것이다.

주요어: 데이터, 표상, 증거적 가치, 데이터 중심 생물학, 레오넬리