The Philosophy of Data

Sabina Leonelli, s.leonelli@exeter.ac.uk

Introduction

In contemporary scientific discussions of big data and data-intensive research, the term "data" is sometimes used to indicate basic, incontrovertible facts on a given entity or process, which can therefore be assumed to provide reliable information about it. This chapter uses key literature in the Anglo-American philosophy of science to show that this conceptualisation of data and their role as evidence is far too simplistic. Philosophers have long been aware of the context-dependent nature of data production and, more generally, the unreliability of sensory perception as grounds for knowledge of reality. A good instance of this is Wilfried Sellars' attack on the “myth of the given”, and particularly his claim that “non-propositional items (such as sense data) are epistemically inefficacious and cannot serve as what is given” (Sellars reported in deVries 2014). This mistrust of data as empirical warrants does not, however, clarify the epistemic status of knowledge grounded on the production and
analysis of data, such as, most notably, the knowledge generated through scientific methods. In considering this issue, this chapter reviews the work of philosophers who have considered the tensions and problems involved in using data as evidence for scientific claims, and the implications that this has for a broader conceptualisation of the nature and function of data. There are of course many ways to think about data that do not pertain to the realm of science. Nevertheless, the production and use of data to produce scientific knowledge is a relatively well-demarcated domain of human activity, and thus constitutes a good starting point for reflections over the status and nature of data in a more general sense. Within the sciences, it is clear that data are deeply historical entities, which are generated (in the case of experimental data) or collected (in the case of data derived from fieldwork) under controlled circumstances to serve as evidence for knowledge claims. Thinking about scientific data can therefore encourage philosophers to avoid ahistorical, uncontextualised approaches to questions of evidence, and instead consider data as components of specific processes of knowledge-making.

**Data in the Philosophy of Science**

Data can be easily construed as a starting point for scientific reasoning about the world, its structure, and functioning. They are the facts from which reasoning proceeds, and the empirical basis for testing and validating any assertion made by scientists about the nature of reality. Within an experimental setting, data are commonly identified with the immediate traces left by measurement instruments and the manipulation of samples and, as such, they are taken to
document features and attributes of the entities or processes under investigation. This is where the idea of “raw data” comes from. Data are as close as a scientist gets to documenting specific aspects of a phenomenon of interest in a way that can inform further inquiry, without necessarily attempting to reproduce or represent the phenomenon itself. They are “raw” because they have not yet been subjected to extensive research interventions, such as modelling and statistical analysis.

Interpreting the scientific meaning of data is left to the researchers who handle them, who decide whether to regard them as evidence for specific phenomena on the basis of their interests, background knowledge, and familiarity with the procedures through which the data were obtained. The importance of human agency in attributing meaning to scientific data provides a starting point for philosophical analysis. For centuries, philosophers have observed that despite their epistemic value as “given”, data are clearly made. They are the results of complex processes of interaction between researchers and the world, which typically happen with the help of complex interfaces such as observational techniques, registration and measurement devices, and the re-scaling and manipulation of objects of inquiry for the purposes of making them amenable to investigation. For example, data about an organism of interest to biologists are usually gathered through the use of instruments such as microscopes, mass spectrometers, genome sequencers; techniques such as control trials and mutant screens; and the modelling of the organism itself in a variety of ways, ranging from the standardisation of its environment (from the field to a laboratory cage) to surgical interventions and the genetic manipulation of its offspring. These
experimental processes embody specific interpretations of the world. For instance, genome sequencing machines incorporate assumptions about how a genetic sequence is assembled and what role it plays within the wider organism, while the ways in which laboratory organisms are kept and fed reflect researchers’ ideas about what constitutes “optimal” nutrition for “normal” development.

Ronald Giere has discussed the large amount of conceptual and material scaffolding involved in scientific data production as exemplifying the perspectival nature of observation – that is, the extent to which what is perceived as a laboratory finding is actually the result of looking at the world through a specific theoretical perspective, honed through years of research, and steeped into well-entrenched assumptions and commitments about how the world works (Giere 2006; see also Gooding 1990). This means that experimental findings are never pristine, objective documents of a mind-independent reality ("raw" in the sense discussed above), but rather the results of situated attempts to interact with the world on the basis of a given worldview. The epistemic significance of data needs to be evaluated accordingly. As claimed by several philosophers and philosophically minded historians and sociologists, this is also the case for data generated outside the controlled environment of the laboratory, such as observations made during fieldwork (see for instance Hanson 1958, Latour and Woolgar 1979, Hacking 1983, Collins 1985, Franklin 1986, Galison 1987, Bogen and Woodward 1988). Building on this work, Hans Radder has argued that field observations are as context- and subject-dependent as experimental results (Radder 2006).
The tension between viewing data as instances of the world and emphasising their man-made nature has acted as a thread for philosophical discussions of scientific methods at least since the scientific revolution. For the most part, philosophers have focused their efforts towards debunking the myth of data as given rather than made. Accordingly, almost every prominent philosopher in the Western tradition has been suspicious of the so-called “method of induction”, which is grounded on the idea that claims about the world may be generated through the accumulation of facts and the emergence of meaningful patterns from such facts. Many have viewed reliance on induction as equivalent to accepting that there can be a set of observations so reliable and fact-like that one can infer truthful generalisations about the world from them – and found this assumption wanting. In his seminal *Essay Concerning Human Understanding* (1690), John Locke noted that humans are far too conditioned by their own assumptions and interests to be able to observe the world objectively. Subsequent scholars have shown scientists to be no exception. Accordingly, scientific methods have been portrayed as efficient means to moderate, and where possible annihilate, such subjectivity – an achievement that presupposes the recognition that what one takes to be a fact about the world may well be a fallacious impression generated by the senses. Whatever fact science proceeds from, one must instead consider how belief in that fact was generated in the first place. In the case of data, this means questioning which instruments, procedures, materials and conceptual assumptions were made in order to produce a given dataset, and evaluate any possible interpretation of data against this background. A particularly vocal advocate of this view was Pierre Duhem, whose 1906 treaty on the structure of physical theory inspired later authors, such as Norwood
Russell Hanson and Thomas Kuhn, to emphasise the inevitable influence of theoretical presuppositions on data collection, selection and interpretation, which they referred to as theory-ladenness (Duhem 1906, Bogen 2010, Schindler 2013).

The theory-laden, man-made nature of data caused much concern within 20th century philosophy of science, because it makes it difficult to think of data as providing objective evidence for given theories. Largely thanks to the influence of logical positivism and Karl Popper’s falsificationism, most Anglo-American philosophers writing after World War II conceived of data chiefly as means to test theories. Within this tradition, data need to provide a benchmark as ‘hard facts’ that can confirm as well as disqualify researchers’ theoretical hypotheses: they are the ground on which theories are validated, and thus need to be reliable and trustworthy. This requirement does not fit well with the realization that data are, at least in part, a reflection of scientists’ specific interests, background knowledge, location, instruments, and research strategies. It also runs against the insight that “publicly available data typically cannot be produced except through processes whose results reflect the influence of causal factors that are too numerous, too different in kind, and too irregular in behaviour for any single theory to account for them” (Bogen 2010, 18). Carl Hempel is one of many philosophers struggling to reconcile the local, idiosyncratic and theory-laden nature of data and their function as conformation for universal truths about nature. His solution was to rely on scientific methods to filter researchers’ “sensations, perceptions and similar phenomena of immediate experience” out of the process of inquiry, leaving only “directly observable” and “intersubjectively
ascertainable” observations that can be taken as objective facts about the world and used to validate a given theory or explanation (Hempel 1970, 674).

**Data in Scientific Practice**

Philosophers’ long struggle with the relation between theory and evidence contributed to establishing a view of scientific knowledge as a set of universally valid claims about the world. Within these accounts, the theory-ladeness of data is problematic because of the perceived tension between the context-dependent nature of data production and the use of data as evidence for supposedly objective, context-independent statements (often referred to as ‘laws of nature’), the discovery of which is the ultimate goal of the science. Theories and explanations that abstract as much as possible from fallible human perception are conceived as the main product of research, and hence as the scientific elements that are most deserving of philosophical scrutiny. As a result, philosophers focused their attention on theoretical debates within the sciences, rather than on observational or experimental practices involving human subjectivity and experience. Following in this vein, the theory-ladenness of data, and hence their embeddedness in specific histories of inquiry, was presented as a threat to the legitimacy of scientific knowledge as a reliable source of insight about the world. It is no wonder that Hans Reichenbach (1938) characterised the messy and sometimes serendipitous processes of data handling as part of the ‘context of discovery’, which he carefully distinguished from the rational marshalling of data into evidence within neat arguments that is involved in the production of scientific claims about the world, a process which he dubbed the
‘context of justification’. In Reichenbach’s view, the value of research as a harbinger of truth is found by scrutinizing how scientists construe and present their conclusions, rather than through an examination of the conceptual and practical constraints that go into producing the data used as evidence for those claims.

This view of scientific knowledge has been challenged by the recent ‘practice turn’ within the philosophy of science. Starting from the 1970s, an increasing number of philosophers started to pay more attention to authors such as Francis Bacon (1994), William Whewell (1989) and John Stuart Mill (1843), who emphasised the fruitfulness of examining the actual features of processes of discovery, rather than their glorified *post facto* reconstruction. This interest has been primarily channelled in the study of the role of scientific models, whose epistemic role has been found to vary depending on their concrete features, ranging from mathematical formalisations to material objects, as well as the interests and values of their users – thus demonstrating the philosophical import of studying actual research practices in detail.1 In response to this scholarship, some philosophers have become convinced that understanding the nature of knowledge and scientific reasoning meant studying the history and characteristics of research practices across different periods, locations and disciplines, including the specific constraints and the variety of worldviews underlying and shaping the production of data and its use as evidence for claims (Ankeny et al. 2011; Chang 2004). A parallel realization has been that science is an exceptionally diverse enterprise, which might be better investigated starting

---

1 Notable contributions to this scholarship include Nancy Cartwright (1983), Mary Morgan and Margaret Morrison (1999) and Ronald Giere (2006), among many others.
from the idiosyncrasies of specific cases rather than from an emphasis on common strategies and overarching theories (Kellert, Waters and Longino 2006; Chang 2012).

In this context, many discussions of relevance to data have centred on “models of data”, i.e. manipulations of experimental data aimed at eliminating errors and producing statistical patterns, which can then be used to test theoretical predictions (Suppes 1962). The idea of models of data helps to understand how scientists transform a set of scattered data points into a smooth curve on a graph, which fits data points onto what scientists take to be a significant pattern.

In Roman Frigg’s and Stephen Hartmann’s contemporary reformulation, models of data are a “corrected, rectified, regimented and in many instances idealized version of the data we gain from immediate observation, the so-called raw data” (Frigg and Hartmann 2012). Focusing on these models is an excellent way to emphasise the complex processes through which data produced by a given set of instruments and/or procedures are marshalled into evidence for specific claims, and particularly the numerous assumptions and constrains underlying not only the production, but also the dissemination and use of data in scientific research.

Another set of discussions in which philosophical attention has turned to data in scientific practice is the philosophy of experiments. Within that realm, a central contribution is that of Ian Hacking, who coined a broad definition of experimental data as *marks* produced by human interactions with research instruments. By focusing on the material circumstances in which data are generated, Hacking’s account remains agnostic about the epistemic role that data may play in scientific inquiry, and indeed does not even require data to function
as evidence for claims about phenomena, though this is of course what data are typically used for in research. Hacking’s objective is instead to stress the constraints and opportunities provided by the manifold of formats and shapes in which data are produced in the laboratory – comprising, in his words, “uninterpreted inscriptions, graphs recording variation over time, photographs, tables, displays” (Hacking 1992, 48). Peter Galison has taken a similar position with respect to data obtained through experiments in particle physics, which has enabled him to study how data are exchanged across research communities in this area and how their scientific use is affected by their movements (Galison 1997).

Hacking’s work inspired James Bogen’s and James Woodward’s seminal account of the relationship between data production and the development of claims about phenomena, according to which data cannot provide evidence for theories, but rather provide evidence for the identification and characterisation of phenomena such as the melting point of lead or the existence of weak neutral currents. It is these phenomena, rather than data, that feature as evidence within theories. One of the key achievements of their approach, in which philosophical analysis is again tightly intertwined with an examination of research processes carried out in experimental physics, has been to resurrect a conception of data as things that can be straightforwardly observed. As they put it, “we need to distinguish what theories explain (phenomena or facts about phenomena) from what is uncontroversially observable (data)” (1988, 314). Bogen and Woodward embrace the fact that data are “typically the result of complex interactions among a large number of disparate causal factors which are idiosyncratic to a
particular experimental situation”, but do not view this as a threat to the potential value of data as evidence. To the contrary, they welcome the study of the messy context of discovery as a crucial starting point for understanding and evaluating how, when and why data can and do function as evidence for specific claims.

Another seminal figure to take inspiration from Hacking’s work, as well as the oeuvre of French philosophers such as Gaston Bachelard and George Canguilhem, is Hans-Jörg Rheinberger. In his account, data are things that can be stored and retrieved, and are thus made durable – a very important characteristic, as data need to be passed around and scrutinized by peers in order to document the claims for which they are presented as evidence (2011). Despite their common emphasis on experimental data practices, Rheinberger’s conclusions differ from Hacking’s insofar as he does not view the marks produced by scientific instruments – which he calls “traces” or “signals” – as an example of data. Rather, he conceives of data as the result of further manipulations of the traces resulting from observation or experiments – manipulations that are performed with the purpose of storing those traces and making them available and intelligible to others. As an example, Rheinberger points to the first DNA sequence gel produced by Fred Sanger and collaborators in 1977. The gel helps to visualise the relatively simple molecular structure of the DNA sequence of bacteriophage PhiX174, by generating discrete stripes of varying lengths on a photosensitive plate. Each stripe is then made to correspond to one of the four nucleic acid bases of DNA (guanine, adenine, thymine and cytosine). Finally, the initial letters of these acid bases (GATC) are
used as symbolic stand-bys for the stripes themselves – an important move, because these symbols can be digitalised and analysed much more easily than the cumbersome and idiosyncratic stripes initially generated by the DNA sequence gel. Rheinberger interprets those stripes as *traces* generated by this laboratory technique: the immediate products of experimentation, which however is difficult to move around in its ‘raw’ state. He contrasts these traces with the abstraction of these stripes into letters that can be easily moved around and used for further research, and refers to these letters as an example of the transformation of traces to *data* (ibid., 6-7). This account benefits from the extraordinary success of the use of letters as symbols for nucleobases, whose format has certainly facilitated the implementation of the molecular bandwagon in biology. Rheinberger also explicitly builds on Bruno Latour’s work on scientific knowledge production, and particularly his analysis of chains of inference. As Latour demonstrates by following the stages through which data have been collected and mobilised to document the botanical and geological characteristics of a given area of the Amazons, the establishment of knowledge claims is grounded in the production and movement of objects that can serve as anchors for knowledge claims thanks to their stability across contexts – and which Latour calls, with characteristic flair, “immutable mobiles” (Latour 1999).

Both Latour and Rheinberger recognise that the marks (or traces) produced in the course of research need to be processed in order to travel, and that travelling across labs and research contexts is crucial to their functioning as evidence. They also emphasise the epistemic importance of the mobility of data and the labour required to realise it. This shifts the philosophical focus from the logical links
between data and claims – which can be analysed without taking any contextual element into account, as characteristic of traditional approaches within the philosophy of science – to the relation between researchers’ perceptions of what counts as data and the contexts in which such perceptions emerge. Latour’s and Rheinberger’s accounts of data are thus strongly influenced by their interest in how scientific practices actually unfold and generate knowledge claims.

**Two Conceptions of Data: Relational and Representational**

In (Leonelli 2009, 2014) one finds an investigation of how data are mobilised and manipulated in order to expand their evidential value, for instance when devising databases capable of making data usable as evidence for a variety of different claims (Leonelli 2009, 2014). As a result, a *relational* account of data is advocated, where what counts as data depends on who uses them, how and for which purposes: “any object can be considered as a datum as long as (1) it is treated as potential evidence for one or more claims about phenomena, and (2) it is possible to circulate it among individuals” (Leonelli, 2015). This account makes two key assumptions about data: that they are portable and that they consist of material objects. Portability is important because the establishment of scientific claims is widely recognized as a social activity that needs to involve more than one individual. Sharing data among individuals can therefore be viewed as a necessary, though not sufficient, condition for their prospective use as evidence. If data are not portable, it is not possible to pass them around a group of individuals who can review their significance and bear witness to their scientific value. Lorraine Daston and Peter Galison (1992) make the same point.
when describing data as quintessentially workable and “communal”, a point extended by Mary Morgan, who stressed the crucial importance of movement across contexts to assessing the value of data as evidence (2012). Materiality is then crucial to making data portable. As also emphasized by Hacking, whether we are dealing with symbols, numbers, photographs or specimens, all data types need a concrete medium in which they can be disseminated. This concrete medium can encompass both digital and analog objects and processes, as long as it is physically possible to pass it around among individuals. The approach follows the lead of Orlin Vakerelov, who defines a medium as “the concrete stuff. It is the system that gets pushed and pulled by the rest of the world […] in just the right way to support the patterns of interactions that constitute the information process” (Vakerelov 2012, 49).

According to the relational view, scientific data can thus be usefully characterised as objects that are explicitly collected and disseminated in order to provide evidence for claims about reality. This does not mean that whoever gathers data already knows how they might be used. Rather, what matters is that data are collected with the expectation that they may be used as evidence for one or more claims about the world at some point in the future. This implies that the same objects may or may not be functioning as data, depending on which role they are made to play in scientific inquiry. This accounts for the diversity of formats, media and context that data typically inhabit in scientific practice: within the relational account, the same objects can change some of their material features and yet be used as ‘the same data’ across a variety of contexts, as well as
cease to function as data as soon as they are no longer regarded as sources of evidence for a claim.

The view contrasts sharply with those of philosophers and scientists who prefer a context-independent definition of what data actually are, a view that can be here broadly characterized as the *representation*al account of data. Within this view, data can be identified regardless of the ways in which they are used at any point in time, and it is possible to evaluate objectively, without any reference to the relevant research context, what information a given dataset contains, and whether this is being interpreted correctly or incorrectly. In a recent report, for instance, the Royal Society proposed to define data as “numbers, characters or images that designate an attribute of a phenomenon” (Royal Society 2012, 12). This definition can easily be interpreted to depict data as representations of a given entity or process, in the sense of providing access to one or more of its characteristics. This reflects the common intuition that data, especially when they come in the form of images like photographs, somehow mirror the phenomena that they are created to document, providing a snapshot of those phenomena that is amenable to study under the controlled conditions of research. It also reflects the idea of data as ‘raw’ products of research, which are as close as it gets to unmediated knowledge of reality. This is a useful view insofar as it makes sense of the truth-value sometimes assigned to data as irrefutable sources of evidence – the Popperian idea that if data are found to support a given claim, then that claim is corroborated as true at least as long as no other data are found to disprove it. As soon as a Popperian view of scientific
progress is abandoned, however, the representational view of data runs into problems.

One problem is that this view makes it difficult to accommodate the wide variety of uses and media for data found in scientific practice. In particular, it restricts any given dataset to the role of representing one (and only one) phenomenon, while, as amply demonstrated by the recent emphasis on re-purposing existing data through making them widely accessible, the same dataset may well be interpreted as representing a variety of phenomena, depending on the expertise and interests of the researchers involved and the level of abstraction (see chapter seven). Another problem with this account is that it makes it hard to account for situations where scientists produce data without knowing exactly what it is that those data are documenting – which attribute of which phenomenon is being represented. Such a situation may emerge, for instance, in the research approach that Friedrich Steinle (1997), Richard Burian (2007) and Maureen O'Malley (2008) have described as exploratory experimentation, where data production is driven by the availability of specific instruments or procedures (such as a genome sequencer that makes it easy to produce sequencing data from any organic sample), and by the hope that data generated through those means will inspire new observations, questions or insights on as yet unknown phenomena. Researchers do not typically start their inquiries with a clear idea of what their data may represent; and even by the end of their study, they may have diverging interpretations of exactly which phenomena are being captured by the data. Perhaps the most famous example of such a situation consists of the photographs of DNA structure made by Rosalind Franklin in the
early 1950s. Franklin, a crystallographer by training, was producing those images to explore the arrangement of nucleic acids in chromosomes. James Watson and Francis Crick, with backgrounds in biology and biochemistry, interpreted the same photographs as evidence for how DNA coding works – a discovery that eventually earned them the Nobel prize. Rosalind’s interpretation of the significance of her data was not wrong nor uninteresting, indeed it foreshadowed other developments in biology, such as the discovery of copying mechanisms for DNA. Nevertheless, Watson’s and Crick’s interpretation was widely regarded as much more important, and this was in part due to their own ambitious quest of finding the mechanism for inheritance (the “code of life”): when they saw the data, they were immediately able to use them as evidence for their hypothesis (in James Watson’s words, “everything fell into place”; Watson 1968).

The relational view on data can make sense of cases like this, but it is not without its problems either. For example, the representational view of data makes better sense of the idea that a specific dataset remains the same even when it is copied in multiple versions or when it changes format, e.g. from a .jpeg to a .pdf file or from one inscription to another (as in the case of the abstraction of stripes into symbols used by Rheinberger, discussed above). In those cases, it is the representational value of the data that defines their identify and continuity in time and space, rather than the specific embodiment of the data at any given moment. Another problem with the relational approach concerns how portability itself is defined, and whether it always requires that data are conceptualised as material entities. Would a sighting of an object by witnesses
count as making that object portable, because the witnesses carry the image of the object in their head? So, when a scientist sees an event that no one else witnesses, which leaves a trace in her memory and which she then tells others about, does this count as data? Such a case is not typically regarded as an instance of data production by today’s scientific institutions (see Steve Shapin’s 1994 history of witnessing in science), and yet this question becomes highly relevant when going beyond the scientific domain and thinking about data in other knowledge domains, such as legal cases.

**Data as Sources of Information**

This chapter has focused on data as artefacts that are taken to carry information about the world, and on the processes through which the attribution of information content to data can be made, from the point of view of the philosophy of science. These issues have strong bearing on the philosophy of information. Despite his insistence on a mathematical interpretation of information, Claude Shannon himself recognized the polymorphic nature of the notion of information and its inter-dependence with fields of application: “It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field” (Shannon 1993, 180). Few philosophers, however, have ventured to examine how data and information are treated within scientific practices, and with which implications. Luciano Floridi has paved the way towards such investigations by providing a framework that places the study of data at the heart of the philosophy of information (see chapters 6 and 7 of this book). In his view, “there can be no
information without data representation” (or “physical implementation”; Floridi 2005). Data thus function both as sources from which information can be obtained and as media in which information can be inscribed. Indeed, Floridi defines information itself as “data + meaning”, and a datum consists of “x being different from y, where the x and the y are two uninterpreted variables and the domain is left open to further interpretation” (Floridi 2011: 85; chapter 6).

Floridi calls this latter characteristic an “uninterpreted difference”, thus stressing the materiality of data in ways that resonate with Ian Hacking’s views on “marks” and more generally with the relational account. This is particularly evident in Floridi’s definition of data as “relata” and “diaphora de re, that is, as lack of uniformity in the real world out there” (ibid., 85-87). Moreover, he provides a taxonomy of data types which is extremely useful when analysing the status and role of data in scientific practice. For instance, he distinguishes between primary data, in which information is encoded, and secondary data, which provide information through their absence, and meta-data, which provide information about the ways in which data came to be, which is highly relevant to interpreting their significance.

One crucial issue raised by Floridi’s interpretation concerns the relation between data and truth-value. Floridi maintains that the truthfulness of data needs to be evaluated in relation to the context in which they are produced, and specifically the goal and level of abstraction for which data are taken to carry information. Thus, truthfulness is not an intrinsic property of data themselves, but rather is determined by the situation in which data are generated. Thus, data are relata with respect to the context in which they have been produced, but their
significance does not change in relation to the contexts in which they are used as evidence – researchers need to uncover the truth-value of data by identifying what they were originally taken to indicate. This position places Floridi’s account beyond accusations of relativism. However, it also moves him away from a more radically relational view, within which the significance of data – the information that they are taken to carry – can vary depending on the context in which they are adopted and used. In this latter approach, the conditions under which truth is established are dependent on the theoretical, material and social commitments of the group(s) involved in interpreting data – a position that can more easily be interpreted as a form of relativism, or ‘perspectivalism’ in Giere’s words. Data are functional components of a process of inquiry, defined by their role as evidence. It is therefore not by looking at data in isolation that questions of truth-value can be addressed; rather, it is by exploring the ways in which data are situated in specific contexts of inquiry.

An advantage of latter approach is its ability to reconcile the *prima facie* contradictory perceptions of data as ‘given’ and ‘made’ which, as noted at the start of this chapter, have long plagued philosophical and scientific discussions of data processing and interpretation. Data do not need to be conceptualised as objective and context-independent units in order to make sense of their value as evidence towards scientific claims. Acknowledging the subjective, context-dependent nature of data is a fruitful starting point when fostering their adoption as evidence for a variety of different claims. This framework also helps to understand the link between data and scientific knowledge production. Data are not, by themselves, a form of knowledge. Rather, data need to be interpreted
in order to yield knowledge; and interpretation, in whichever form and through whichever process it is achieved, involves using data as evidence for one or more claims about phenomena. It is those claims, rather than the data, which express knowledge about reality, and are therefore often referred to as 'knowledge claims' or *propositional knowledge*. This form of knowledge is also what scientists typically refer to as expressing the scientific significance of data.

In closing, a few words about the notion of evidence used in this chapter. A feature that all the above accounts of data have in common is an understanding of evidence as the grounds on which specific claims about reality acquire credibility. In more general terms, evidence is here assumed to consist in whatever makes a given assertion believable, or anyhow increases its intelligibility and/or plausibility to a given audience. The definition is necessarily broad, as there are several ways in which evidence may be provided, not all of which are in the form of data. For example, a logically constructed argument may function as evidence, by increasing the plausibility of a given claim (as when I tell my toddler son that “lying is dangerous”, and provide evidence by arguing that “it is very hard to hide the truth, and if you are found out, you may be punished”). Specific actions, such as pointing to an object or staging a demonstration, can also function as evidence for assertions (e.g. supporting the claim “the leaves have fallen from the trees” by pointing to a nearby forest, or the claim “it is possible for adult humans to jump over this fence” by successfully doing it). This definition of evidence restricts the notion to a relation with a propositional statement, which may be overly restrictive when considering situations where

---

2 An overall review of philosophical debates around the notion of evidence is provided by Thomas Kelly (2014).
evidence is accrued to demonstrate a state of affairs not easily captured by language. Whether the epistemic role of data as evidence is always in relation to a proposition is a question worth asking, though existing accounts of the epistemic status of data tend to give this for granted.

References


**Biographical Note**

Sabina Leonelli is Associate Professor in Philosophy and History of Science and Associate Director of the Centre for the Study of the Life Sciences (Egenis) at the University of Exeter. She uses empirical research to foster philosophical understandings of knowledge-making practices.