Empirical Underdetermination: The Empirical Side of the Duhem-Quine Thesis

María Caamaño

Philosophy Compass **2025**

Proyecto PID2021-126416NB-I00 financiado por MICIU/AEI /10.13039/501100011033 y por FEDER, UE

Abstract

Theoretical underdetermination is a central issue in the Philosophy of Science, having been discussed and debated since the early 20th century. The so-called "Duhem-Quine problem" has been used as an umbrella term to refer to a number of problematic features that arise from the lack of a biunivocal correspondence between theory and evidence. However, the now familiar idea that the detection of an empirical phenomenon is inferred from a complex collection of data (Bogen & Woodward 1988, Woodward 1989, 2000, 2010, McAllister 1997, 2011, Glymour 2000, Harris 2003, Massimi 2007, Leonelli 2015, 2019, Bokulich 2020) entails the recognition that not only theories, but also the description of empirical phenomena are underdetermined by evidence. Empirical underdetermination, understood as the underdetermination of evidence (or assumed empirical phenomena) by data, emerges as a major challenge that has yet to be fully recognized and carefully addressed in the philosophy of science. The paper summarizes the distinction between empirical and theoretical underdetermination as it implicitly appears in the literature to date. It presents them as instances of a more general type, both of which arise from the same basic problems, albeit at different levels and with different implications. Important but often overlooked aspects of the empirical/theoretical distinction, the notion of background assumption, and the different roles of evidence will be clarified.

Keywords: underdetermination, evidence, Duhem-Quine thesis, holism, ampliative inference, background assumptions.

Introduction

The starting point for this analysis is the recognition that underdetermination extends beyond the relationship between theories and evidence to include the relationship between evidence (or supposed empirical phenomena) and data. The latter extension of underdetermination is what I call here ‘empirical underdetermination’, to make a difference compared to ‘theoretical underdetermination’. The main goal of this paper is to provide an explicit and systematic characterization of empirical underdetermination. In order to achieve this goal, I summarize the distinction between empirical and theoretical underdetermination as it implicitly appears in the literature to date. Both forms of underdetermination are presented as instances of a more general type, both arise from the same basic problems, albeit at different levels and with different implications. Significant efforts to understand the production of data models in specific contexts of scientific inquiry have been made in recent decades. Most often, the goal of such efforts (Kaiser 1991, Leonelli 2009, Karaca 2018, Bokulich & Parker 2021, Antoniou 2021) has been to characterize the inference from data to empirical phenomena that are potentially useful as evidence. Despite their relevance, these case studies have not included a clear recognition and discussion of the phenomenon of underdetermination at the empirical level. As a result, the implications of these studies for empirical underdetermination are still to be seen. This paper addresses what is missing in the first place: a clear and systematic characterization of empirical underdetermination.

In what follows, I will show what theoretical and empirical underdetermination have in common and what differentiates them, by systematizing the main contributions available on the subject. To do so, I first recall the main features and types of theoretical underdetermination (section 1). I then show how to move from theoretical to empirical underdetermination. In section 2, after making some necessary clarifications, I discuss how empirical underdetermination resembles theoretical underdetermination in both its causes and its forms. Finally, I emphasize the methodological importance of the separate assessment of underdetermination at each level.

1. From theoretical to empirical underdetermination

In the philosophy of science, debates about underdetermination have mainly revolved around theoretical underdetermination, usually understood as the possible or actual coexistence of alternative theoretical explanations given the available evidence. The intuition behind the recognition of this problem is that theoretical explanations of what we observe necessarily go beyond what we observe, otherwise the observation would be transparent, self-explanatory as to its causes and nature. But this is not the case. So we have to develop conjectures or theories that go beyond the observed. Now we can do this in different ways, because the available set of observations (the explanandum) is always compatible with more than one theoretical explanation (the explanans). Moreover, depending on the background assumptions made, the same theoretical explanation may or may not be compatible with the same set of observations.

The discussion of this issue has a long history, which I will not recount here. [1] Rather, in order to show that very similar features can be seen in what I will call empirical underdetermination, I will focus on some central features attributed to theoretical underdetermination. My claim is that, because the empirical domain is complex and multilayered, it also involves a relationship between an explanandum in the form of data and an explanans in the form of assumed empirical phenomena. Not only theories, but also evidence (assumed empirical phenomena or models of data) are vulnerable to underdetermination. In those contexts or fields of inquiry where the assumptions underlying observations are highly conjectural or poorly developed, this problem is particularly challenging.

1.1. The standard approach to underdetermination

The problem of underdetermination is usually referred to as the "Duhem-Quine problem". This is an ambiguous label for two different kinds of underdetermination, namely the holistic and the contrastive, corresponding respectively to the problems of confirmation holism and empirical equivalence (or confirmatory equality) of alternative theories (Newton-Smith 2001, Stanford, 2021). In its holistic version (Duhem 1906/1991, Quine 1951, Okasha 2002, Dietrich & Honenberger 2020), a theoretical hypothesis is always underdetermined by evidence in the sense that it cannot be empirically tested in isolation from a given set of observations. To be testable, a hypothesis must always be conjoined with background assumptions. However, the same hypothesis and initial set of observations are in principle compatible with different sets of background assumptions about the world, the functioning of

[1] See Bonk 2008, Biddle 2013, Turnbull 2017, Stanford, 2021, for systematic accounts of this issue.

instruments, etc. Consequently, the same set of observations may or may not support the hypothesis, depending on the choice of background assumptions.

In its contrastive version (Quine 1975, van Fraassen 1980, 1983, Sklar 1981, Newton-Smith 2001, Bonk 2008, Godfrey-Smith 2008, Lyre 2011, 2018, Acuña & Dieks 2014, Stanford 2021), the available evidence is never sufficient to determine the truth of one theoretical hypothesis over another, provided that both are empirically adequate. Taken together, the two versions of underdetermination entail a lack of a biunivocal correspondence between a given hypothetical assumption and a single set of empirical indicators that provide the empirical evidence to support it (see table 1 at the end of the paper). For clarity, it is useful to further distinguish, from the beginning, between the traditional sense of contrastive underdetermination, which essentially involves empirical equivalence, and what has been called 'transient underdetermination' (Sklar 1975, 1981) or 'practical underdetermination' (Biddle 2013, Turnbull 2017), which refers to alternative theories that─at least for a certain period of time─are equally well confirmed, but not empirically equivalent. However, it seems preferable to call the second type of contrastive underdetermination 'confirmatory equality', as opposed to 'empirical equivalence' underdetermination, rather than Sklar's terminology, since in principle any form of underdetermination can be transient in a literal sense.

Pierre Duhem (1906/1991) convincingly argued that it is impossible to test a hypothesis in isolation. In order to derive empirical consequences from a hypothesis, it must be combined with many other assumptions and hypotheses about the world, the functioning of measuring instruments, environmental conditions, etc. These holistic features of confirmation lead to the recognition of holistic underdetermination. In principle, there are several possible choices of auxiliary assumptions to be conjoined with a hypothesis. Duhem's 1906 paper provides a classic example of holistic underdetermination. As he points out, when testing thermodynamic hypotheses, we need to be able to empirically determine changes in temperature by correlating them with changes in some other quantity. If we use a mercury thermometer to do this, we have to assume that changes in the length of the mercury strand are relevant to detecting changes in temperature. We also have to make numerous assumptions about how mercury expands or contracts as the temperature rises or falls. According to Duhem, this type of measurement depends on the assumption of certain laws of nature, such as linear expansion, according to which the change in length is directly proportional to the change in temperature. There are also assumptions about the conditions under which a temperature reading from a mercury thermometer should be disregarded. For example, if the mercury thermometer is placed in a strong magnetic field. As is well known, Duhem emphasizes as an important implication of his view that confirmation holism excludes the possibility of carrying out crucial experiments. He thus denies that there was a crucial experiment that led to the rejection of the particle theory of light in favor of the wave theory of light.

Willard Van Orman Quine not only acknowledged confirmation holism (1951), but took Duhem's argument a step further by claiming that a theory can always avoid refutation by changing the auxiliary assumptions conjoined with it (Quine & Ullian 1970).[2] While accepting the fact of confirmation holism, Popper rejected the implications usually drawn from the so-called Duhem-Quine thesis. In particular, he rejected the idea that when a false prediction is derived from a hypothesis combined with auxiliary assumptions, it is not possible to identify where the error lies (Popper 1963, 322-25). Against this "holistic dogma", he claimed that it is always possible to identify the logical connections between hypotheses or assumptions and refuted predictions. The way to do this would be similar to the proof of independence of axioms in formal systems, which would be to find a model that satisfies all axioms except the independent one. If some refuting evidence is gathered, it may be that such evidence provides a model that satisfies several of the assumptions but does not satisfy the main hypothesis that happens to be associated with them. If so, even in non-axiomatized systems, we could identify the source of the error by conjoining another hypothesis to the same assumptions and checking whether the previously refuting evidence is now a model of the new system sharing the same auxiliary assumptions as the old system. In this case, if the result is positive, we have good reason to conclude that the assumptions were not the source of error in the first place when they were in conjunction with the old hypothesis. Interestingly, the Quinean holistic claim that a theory can always be immunized against contrary evidence links both forms of underdetermination. For once a theory has been "immunized" in a holistic way–i.e. by changing some auxiliary assumptions associated with it–there is no conclusive way to discard it in favor of a rival theory, thus favoring contrastive underdetermination. In order to rationally compare rival theories, Quine invokes six pragmatic norms: conservatism, modesty, simplicity, generality, refutability and precision –each of them appealing to a different theoretical virtue (1970/1978, 67-79, 1990/2003, 14–15).

On the other hand, a famous example of contrastive underdetermination was provided by Bas van Fraassen (1980), who described a case where we have two alternative Newtonian cosmologies with the same predictive capacity: one of them would include the Newton's laws of motion, the law of universal gravitation and the assumption that the universe, as a whole, is

[2] Two illuminating analyses of the problem of the scope of confirmation holism are found in Ariew 1984 and Moulines 1986.

stationary; the second cosmology would differ only in that the assumption added to the laws would be the opposite.[3] A peculiar and controversial feature of van Fraassen's example is that the empirical equivalence is the result of an epistemic limitation, and one that seems impossible to overcome. From our position in the universe, we cannot, in principle, detect the constant absolute motion of the universe as a whole. In this case, empirical equivalence would go hand in hand with the inclusion of empirically vacuous assumptions, which for this reason could be regarded as superfluous theoretical components. The same problem arises in the context of van Fraassen's (1983) and Kukla's (1996, 145) use of theory-producing algorithms to generate an empirically equivalent rival theory for every theory in science, as noted by Norton (2008, 27-29).[4]

Broadly speaking, contrastive underdetermination occurs when rival theories are all empirically adequate (Quine 1975). It occurs when alternative theories are highly confirmed empirically, whether or not they are empirically equivalent (Sklar 1981). It has to do with the inability (transient, recurrent or permanent) to reject one theory in favor of another on the basis of evidence.

On the other hand, as several authors have noted (Sklar 1981, Kitcher 1992, Laudan & Leplin 1991),[5] if we assume that it is always possible to gather new evidence sometime in the future, we could consider all cases of contrastive underdetermination transient. By the same token, however, this type of underdetermination could be recurrent (Godfrey-Smith 2008, Acuña & Dieks 2014). Thus, underdetermination can always strike again, even in the form of potential unconceived alternatives ‒that may or may not emerge in the future (Stanford 2021).

---

[3] Darren Belousek (2005, 670) uses the same argument as van Fraassen to argue for the observational indistinguishability of rival theories in quantum mechanics, emphasizing that: "any experimental test that (dis)confirms 'orthodox quantum mechanics' (dis)confirms Bohmian mechanics, and vice versa". Jeremy Butterfield (2012) illustrates the same situation in the case of cosmological models, which have the same empirical consequences.

[4] Norton (2008, 35-38) identifies underdetermination in general with contrastive underdetermination in the form of empirical equivalence. From there he goes on to argue that if it is possible to establish that two theories are observationally equivalent, then, in his own words, 'they will be sufficiently close in theoretical structure that we cannot rule out the possibility that they are merely variant formulations of the same theory'. His objection to this form of underdetermination is very close to the one that has been raised here, for both stress that observationally equivalent theories will most often differ in terms of additional structures that do not represent anything physical and thus have no empirical significance.

[5] A well-known critique of empirical equivalence underdetermination arguments is provided by Larry Laudan and Jarrett Leplin (1991, 451-455, 461-465). They point out that the accumulation of scientific knowledge over time means that some theories often gain evidential support over others. Furthermore, they argue that empirical consequences are not the only source of evidential support for theories.

1.2 Empirical underdetermination

The problem of empirical underdetermination has rarely been explicitly formulated, although it is implicit in various discussions, such as Duhem's holistic thesis, the theory-ladenness of observation, models of data, or incommensurability of experimental practices. Hacking's concept of literal incommensurability (1983, 1992) and Pickering's of mechanical incommensurability (1984, 1995) imply that in laboratory sciences (like particle physics) alternative theories are true for different kinds of phenomena "created" in the laboratory with different kinds of instruments. It is through disjoint sets of measurement procedures that rival theories are, in their view, applied and justified. In addition, the creation of instruments is very often subordinated to theoretical interests. Léna Soler (2008, 327) points to the famous weak neutral current case, where conflicting LIPs [local interpretation procedures] about experimental phenomena are associated with alternative experimental procedures.

A notable step towards the recognition of empirical underdetermination as a distinct form of underdetermination is Thomas Bonnin's (2021) account of 'pervasive underdetermination'. The case study from evolutionary biology that he analyses is understood–as in the case of the weak neutral current–as a direct consequence of methodological incommensurability. He stresses the need to treat evidential claims as evidential hypotheses, thus explicitly acknowledging the conjectural nature of evidence. In his example, we find a theoretical underdetermination involving two rival theories on the origin of eukaryotic cells, the phagotrophic and the syntrophic theories. The first, promoted by Tom Cavalier-Smith, postulates the intervention of phagocytosis, i.e. the ability of cells to engulf and digest other cells. On the other hand, Bill Martin's syntrophic theory postulates the action of metabolic approximation, in other words, the intervention of a hydrogen-consuming host that would progressively surround a hydrogen-producing alphaproteobacterium. To test these theories, we need to establish the relative timing of the origin of phagocytosis and the origin of mitochondria (Bonnin 2021, 140). But there is an empirical under-determination here, because evolutionary biologists disagree about which of the following evidential claims are true:

(i) Mitochondria were caused by phagocytosis [Tom Cavalier-Smith],

(ii) phagocytosis was caused by mitochondria [Bill Martin].

Although the details of the example are omitted, it illustrates how evidence is constructed in different, incompatible ways from the same pool of available data. This is because Tom Cavalier-Smith and Bill Martin use different data selection, different investigative scaffolds for interpreting the data, and different interpretations of background assumptions. For instance, data about an insect called mealybugs was employed by Martin to defend ii) but Cavalier-Smith argues in favor of i) discarding mealybugs as similar enough to the phenomenon of interest. Another argument by Cavalier-Smith suggests that the external environment is incompatible with the transformations allowed by the cellular environment of the mealybugs. They also have divergent interpretations of background assumptions on different physiological constraints imposed on cells, like the constitution of the endomembrane system, which would be endogenous, according to Martin, and would require other, more energy consuming physiological parameters in the case of Cavalier-Smith (*ibid*., 146).

Woodward (2000) emphasizes the importance of selecting only those accepted claims or background theories that are well established in order to maximize the reliability of inferences from data to phenomena. However, it is not clear that a restriction of the choice of background assumptions in the way he suggests is always feasible or appropriate. As Bonnin's example shows, the recognition of well-established knowledge may not be a trivial matter for scientists. Moreover, the common recognition of such knowledge may not be sufficient to avoid empirical underdetermination, since scientists may still disagree about its implications for a given question. Thus, as in the case analyzed by Bonnin, background knowledge may be used in incompatible ways when drawing inferences from data to phenomena.

The problem of empirical underdetermination also appears in the literature on models of data, initiated by Patrick Suppes (1962), and in the data/phenomenon framework put forward by James Bogen and James Woodward (Bogen & Woodward 1988, Woodward 1989, 2000, 2010). According to Suppes, models of data are usually part of a hierarchy of models. Their role is to link experimental models (used to characterize the procedure for testing theoretical models) to the experimental design and ceteris paribus assumptions about a particular experimental protocol and setting. Moving down the hierarchy therefore includes: primary (or theoretical) models (at the top), experimental models, data models, experimental design and ceteris paribus conditions. Primary models are the means by which we can break down a substantive question into more local questions that are amenable to empirical testing. The purpose of experimental models is both to relate the primary questions to the (canonical) questions about a particular experiment at hand, and to relate the data to those experimental questions. Data models, on the other hand, make it possible to model raw data in order to put them into canonical form, and also to check that the data generation satisfies various assumptions of experimental models–for example, the assumption of homogeneity of control and treatment groups in a randomized comparative experiment. Below the data models are the experimental design models, which specify the planning and procedural steps for generating data. These experimental protocols are crucial for assessing experimental errors or the influence of extraneous factors. At the lowest level of the hierarchy are ceteris paribus assumptions, which concern those factors that are assumed to remain constant or under control and thus not to interfere with the experimental intervention.[6]

The gap between the realm of finite, discrete data collected in the laboratory or the field and the general properties and continuous quantities involved in detecting empirical phenomena - and thus in obtaining evidence for theory - is an aspect highlighted by Suppes that is of great relevance to the issue of empirical underdetermination (Suppes 1962, 254). In other words, the empirical parameters (needed to test the theoretical models) cannot be identified with anything directly available in the data. In Suppes' illustration, changes in learning capacity - essential for testing theoretical models of linear response theory - cannot be equated with any concrete performance in a given time interval or with any collection of such performances. We need to model the data collected in the experiments in order to estimate the value of these parameters. Modeling the data is what makes it possible to go beyond a particular set of data to try to interpret the data as phenomena.

The recognition of an inferential relation from data to empirical phenomena and the idea that such a relation involves a myriad of assumptions, some of which may be problematic and need to be evaluated, are shared by the data/phenomenon framework developed by Bogen and Woodward and Suppes' hierarchy of models framework. Indeed, several authors (Kaiser 1991, Leonelli 2009, Karaca 2018, Bokulich & Parker 2021, Antoniou 2021) have referred to both frameworks when approaching the issue of the relation between data and evidence. These authors' thorough analysis of the processes and assumptions used to generate data models is extremely valuable for understanding how empirical underdetermination can occur. Kaiser's notion of the inference ticket (1991, 122) or Antoniou's idea of auxiliary data models (2021, 8) are cases in point. Both concepts capture the fact that empirical phenomena are established

[6] Both Ronald Laymon (1982, 108-113) and Deborah Mayo (1996, 132-139, Chapter 8) use the example of the experimental tests of the general theory of relativity on the deflection of starlight to illustrate these different levels involved in theory testing. In their example, the deflection of light is the empirical phenomenon that plays the role of evidence.

by inference from data with the help of theoretical knowledge, sometimes obtained in laboratory experiments, sometimes available in different disciplines.

The empirical scope of underdetermination has also been a concern in the philosophy of Helen Longino. According to her approach, which she calls 'contextual empiricism' (1990, 1996, 2002), the gap between hypotheses and data, which can occur at different levels, makes recourse to values inevitable in determining the acceptability of hypotheses. Moreover, value-laden background assumptions would be essential to the enterprise of determining what counts as evidence. Like other feminist-oriented philosophers, she believes that Quine's pragmatic norms (see section 1.1.) fail to recognize the wide variety of non-epistemic values that play an important role in the determination of our scientific beliefs. She explores this issue in an analysis of research in the fields of human evolution and neuroendocrinology (Longino 1990). Her argument begins with the recognition of holistic underdetermination and the semantic gap between descriptive terms used in describing data and in expressing hypotheses, to conclude that value-laden background assumptions mediate the evidential relations between hypothesis and data. In her view, the only way to prevent the arbitrary dominance of subjective preferences is through critical interaction between members of the scientific community or between members of different communities.

From the above we can conclude that it only makes sense to speak of empirical underdetermination if the conjectural and multi-level nature of evidence is acknowledged. Empirical underdetermination presupposes (at least) a distinction within the empirical level between data and empirical phenomena, just as theoretical underdetermination presupposes a distinction between the empirical and theoretical levels. It is important to distinguish one form of underdetermination from the other because they occur at different levels and have different epistemological and methodological implications. We could certainly conflate the two and speak simply of 'theoretical underdetermination' to refer in general to cases where the same evidence (whether data or empirical phenomena) can be explained according to alternative theories (whether low-level or high-level). An all-encompassing/inclusive notion of underdetermination (including both the holistic and contrastive versions) can be characterized by a non-biunivocal relationship between observation and theory, such that: the same set of observations can be explained by alternative theories, and depending on the choice of background assumptions, the same theory may or may not be confirmed/refuted by the same set of observations. However, this way of speaking could lead to misidentifying cases of empirical underdetermination, where the same data can alternatively be understood as being originated by alternative empirical phenomena ‒hence leading to alternative pieces of evidence−, as cases of theoretical underdetermination, where the same empirical phenomena can be interpreted as caused by alternative theoretically postulated phenomena.

But let's come back to the main issue of the distinction between the theoretical and the empirical underdetermination. To take a classic example, we can distinguish between, on the one hand, the transient theoretical underdetermination affecting the Ptolemaic geocentric model and the Copernican heliocentric model, given the common observations at the time when both were rival theoretical models, and, on the other hand, the (transient) empirical underdetermination affecting the determination of the parallax. In this case, the same empirical phenomenon, i.e. the lack of observed parallax, was taken as a refutation of the Copernican model by its opponents, but not by its supporters. Based on the same data about the positions of the fixed stars relative to the observer, the same empirical phenomenon − i.e. the absence of parallax − was considered to be the case or not to be the case, depending on the background assumptions that were made. Because of the uncertainty about the effect of distance on parallax, the empirical phenomenon of parallax was underdetermined with respect to the available observations. In other words, accepting the empirical phenomenon of parallax was underdetermined given the available observations. This is an example of holistic empirical underdetermination. But empirical underdetermination, like theoretical underdetermination, can also occur contrastively: a) different datasets support different inferences about different empirical phenomena, b) the same dataset supports different inferences about different empirical phenomena. In both cases, different explanatory paths are followed. In the first case, however, the differences also affect the choice of the dataset that is taken as evidence for a phenomenon. Contrastive empirical underdetermination (of type a) is illustrated in Bonnin’s example, mentioned earlier.

2. **The parallelism between theoretical and empirical underdetermination**

Let us have a closer look at the relationship between theoretical and empirical underdetermination. First of all, the very distinction between the empirical and the theoretical needs to be clarified. In addition, given its relevance to the issue of underdetermination, the constitutive or non-constitutive role of background assumptions in relation to a theory needs to be considered.

2.1. <u>Clarifying the empirical/theoretical distinction and the role of background assumptions</u>

Let us make an important clarification of the empirical/theoretical distinction before going further into the problem of empirical underdetermination. Two main features are of particular importance in a minimal characterization of the distinction:

1) the functional role in relation to a theory;

2) the epistemological status in relation to a theory.

1) refers to the role played by one component of inquiry with respect to the other. This role is determined by the relation between the *explanans* and the *explanandum*. This is the view of Horwich and Ben-Menahem, who support the idea that the distinction between theory and observation is relative (Okasha 2002, 316). 2) concerns the stronger or weaker conjectural character depending on the role played, a character that depends on the higher or lower conjectural nature of one component of inquiry relative to the other. Usually we expect the two features to be aligned, i.e. the *explanandum* to be less conjectural than the *explanans*. Taken together, both features, 1) and 2), reveal the context-dependence nature of the *explanans*/*explanandum* distinction.

It is important to note that the data/phenomenon distinction shares characteristics 1), 2) with the empirical/theoretical distinction. In other words, data are the observations (E1) that support beliefs about empirical phenomena (T1) (Suppes 1962, Bogen & Woodward 1988, Woodward 1989, 2000, 2010). Consequently, what serves as *explanans* (i.e., theory) in one context may serve as *explanandum* (i.e., describing an empirical phenomenon) in another, and vice versa. For example, a kinematic description may serve as the explanans for the data on different positions of astronomical bodies or as an explanandum for gravitation. Similarly, gravitation can be considered as an explanans for certain kinematic phenomena or as an explanandum for General Relativity.

A further clarification is also necessary, this time in relation to background assumptions. There are two types of background assumptions, depending on how they relate to the primary theory. Some background assumptions are embedded as part of the primary theory, while others are not. In the former case, background assumptions are not variable in relation to the primary theory, whereas in the latter they are. [7] This distinction is therefore important in analyzing holistic underdetermination, because only in the second case could the background assumptions involved in testing the primary theory be replaced by others, leading to different results when testing the same primary theory. In contrast, in the first case, if the background assumptions change, this would entail a change in the empirical meaning of the primary

[7] A similar distinction is drawn in Balashov's (1994) "string" model of scientific tests.

theory, since these assumptions are embedded as part of the empirical meaning of the theory. For example, depending on whether we grant or not that background assumptions about the conditions for observing parallax are embedded in the Copernican system, then the lack of observed parallax at the time could not or could be accommodated in the primary theory. In the Phlogiston Theory, the laws of classical mechanics were embedded in the theory and assumed as a background theory for the measurement of changes in the weight of different portions of substances before and after combustion. In contrast, background assumptions about initial conditions, such as those involved in estimating the effect of heat in the apparatus, were not embedded in the theory, and depending on the choice of such assumptions, anomalous results could or could not be reconciled with the primary theory.

The term "background assumptions" is ambiguous not only in the first sense, that is, as regards the relationship between them and the primary theory, but also in the sense that it can cover both cases of well-established assumptions and cases of not well-established assumptions. This ambiguity has led to confusion in the debate on holistic underdetermination, since not all background assumptions have the "satellite" nature of those that can be replaced without changing the primary theory.

2.2. Tracing the similarity between theoretical and empirical underdetermination.

Both theoretical and empirical underdetermination can occur in a local or global way. In the former case, the same evidence constitutes a common *explanandum* for which different (high-level/low-level) explanations are provided. Empirical equivalence is the empirical relation that characterizes local underdetermination. In the second case, different pieces of evidence constitute non-shared *explananda* for which different (high-level/low-level) explanations are developed.[8] In its global version, confirmatory equality is the empirical relationship that characterizes underdetermination (see table 2 at the end of the paper).

Both theoretical and empirical underdetermination arise from the same basic problems, but they arise at different levels. These problems have to do with the ampliative nature of explanatory inference and the different choices of background assumptions involved in such inference. On the theoretical level, the inference from observations to theoretical

[8] The second case has usually been discussed in the context of methodological incommensurability, like in Hacking's literal incommensurability (1983, 1992) and Pickering's machinic incommensurability (1984, 1995). Earlier, both Kuhn (1962/1970) and Feyerabend (1970) had drawn attention to extreme cases of incommensurability, where alternative (rival) theories are supported on the basis of alternative evidence.

(postulated/conjectured) phenomena explaining the observed is not univocal. On the empirical level, the inference from the data to empirical phenomena explaining such data is not univocal neither. Different empirical phenomena may explain the same data (contrastive underdetermination), and assumptions about the occurrence of particular empirical phenomena may or may not be supported by a given set/collection of data, depending on the choice of background assumptions that play a role in explaining the data on the basis of the assumed empirical phenomenon (holistic underdetermination). In short, the main problem arises from the fact that both inferences are ampliative, i.e. both empirical and theoretical phenomena are postulated to some extent, both are explanatory (albeit at different levels). Moreover, since the choice of assumptions is always multiple, there is (holistic) underdetermination whenever background assumptions are required. If we acknowledge that background theories are involved in the identification of phenomena from a given set of data, then phenomena are underdetermined by data.[9]

3. Conclusions

Underdetermination is an issue at the empirical level of scientific inquiry, not just at the theoretical level. Although underdetermination at both levels has some features in common in terms of the reasons why it occurs and the forms in which it occurs, failure to pay explicit attention to empirical underdetermination could make us oblivious to the problematic, conjectural nature of evidence. Empirical underdetermination is a common situation in scientific inquiry, namely, a situation where a given pool of available data does not determine the inference to a certain description of the empirical phenomenon causing such data. In such cases, the relevant evidence to test rival theories is underdetermined, which may ultimately lead to establishing such evidence in conflicting ways.

On the other hand, the treatment of empirical and theoretical underdetermination as distinguishable issues opens the way for us to draw the appropriate evaluative and methodological implications at each level of inquiry. The inference from data to empirical phenomena - which is always a prerequisite for the inference from empirical phenomena to theory - is worthy of separate consideration.

[9] By the same token, if background theories are involved in the determination of processed data, then processed data are underdetermined by less processed data. This would lead us to an even broader notion of underdetermination which, for reasons of space, will not be discussed here.

If there is holistic and contrastive underdetermination at the empirical level, it is because the functional distinction (*explanans*/*explanandum*) and the epistemological distinction (more/less conjectural) can also be made at this level. Evidential claims are low-level theories (or conjectures) about what empirical phenomenon explains a particular collection of data.

Underdetermination - in both forms and at both levels - is an inevitable feature of scientific claims, in so far as they are concerned with applicability to particular domains and with ampliative reasoning. When applying a theory (low level or high level) to a particular empirical domain, background assumptions are always needed. This is because many details of such a domain that are not considered by the theory have to be established in order to link the theory to the domain. The choice of background assumptions is always multiple. Thus, holistic underdetermination is inevitable too. Similarly, explanatory inferences, whether from data to phenomena (low-level theory) or from phenomena to (high-level) theories, are always ampliative and therefore not determined by the *explananda*.

The above discussion is mainly conceptual and clarifying. However, it has methodological implications for the evaluation of evidence. Assumptions behind observations should not escape scrutiny. Otherwise, transient empirical underdetermination may be prolonged unnecessarily. A milestone in the methodological treatment of underdetermination along the above lines is Mayo's severe testing approach (1996, 2010, 2018). Ultimately, questions relating to the quality of the evidence could be more clearly distinguished from those relating to the quality of the explanation.

References

Antoniou, A. (2021) “What is a data model?” *European Journal Philosophy of Science,* 11 (4): 1-33. https://doi.org/10.1007/s13194-021-00412-2

Ariew, R. (1984) “The Duhem Thesis,” *British Journal for the Philosophy of Science*, 35: 313–25.

Balashov, Y. (1994). “Duhem, Quine, and the multiplicity of scientific tests”. *Philosophy of Science* 61 (4):608-628.

Biddle, J. (2013), “State of the field: Transient underdetermination and values in science”, *Studies in History and Philosophy of Science*, 44: 124–133.

Bogen, J.,and J.Woodward.1988.“Saving the Phenomena.” *Philosophical Review* 97:303– 52.

Bokulich, Alisa (2020). “Towards a Taxonomy of the Model-Ladenness of Data”. *Philosophy of Science*, *87*(5), 793-806. doi:10.1086/710516 .

Bokulich A, Parker W. (2021) Data models, representation and adequacy-for-purpose. Eur J Philos Sci. 2021;11(1):31. doi: 10.1007/s13194-020-00345-2. Epub 2021 Jan 29. PMID: 33584875; PMCID: PMC7846529.

Bonk, Thomas, (2008) *Underdetermination. An Essay on Evidence and the Limits of Natural Knowledge*, Dordrecht: Springer.

Bonnin, T. (2021) “Monist and Pluralist Approaches on Underdetermination: A Case Study in Evolutionary Microbiology”, *Journal for General Philosophy of Science*, 52:135-155.

Dietrich, M., Honenberger, P. (2020) “Duhem's Problem Revisited: Logical versus Epistemic Formulations and Solutions”, *Synthese*, 197: 337–354.

Feyerabend, P. K., (1970). “Against Method: Outline of an Anarchistic Theory of Knowledge”. In M. Radner and S. Winokur (ed.), *Analysis of Theories and Methods of Physics and Psychology*, (Minnesota Studies in the Philosophy of Science, Volume IV), Minneapolis: University of Minneapolis Press: 17–130.

Franklin, A. (1986). *The Neglect of Experiment*. Cambridge: Cambridge University Press.

Godfrey-Smith, P. (2008). “Recurrent transient underdetermination and the glass half full”. *Philosophical Studies* 137 (1):141 - 148.

Hacking, I. (1983). *Representing and Intervening. Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hacking, I. (1992). “The Self-Vindication of the Laboratory Sciences”, *Science as Practice and Culture*, A. Pickering (ed.). Chicago, University of Chicago Press: 29–64.

Kaiser, M. (1991). “From rocks to graphs — the shaping of phenomena”. *Synthese*, 89 (1): 111 - 133.

Karaca, K. (2018). “Lessons from the Large Hadron Collider for model-based experimentation: the concept of a model of data acquisition and the scope of the hierarchy of models”. *Synthese*, *195*(12), 5431–5452.

Kuhn, T. S., (1962-1970). *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Laymon, R. (1982) "Scientific Realism and the Hierarchical Counterfactual Path from Data to Theory", *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1: 107-121.

Leonelli 2009. "On the Locality of Data and Claims about Phenomena." *Philosophy of Science* 76 (Proceedings): 737–49.

Leonelli (2015) "What count as scientific data? A relational framework" *Philosophy of Science*, Vol. 82, No. 5 (December 2015), pp. 810-821

Leonelli (2019) "What distinguishes data from models?" *European Journal for Philosophy of Science* (2019) 9 (22). https://doi.org/10.1007/s13194-018-0246-0

Longino, H. (1990). *Science as Social Knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.

Longino, H. (1996). "Cognitive and non-cognitive values in science: Rethinking the dichotomy". In L. Hankinson Nelson, & J. Nelson (Eds.), *Feminism, Science, and the Philosophy of Science* (pp. 39–58). Dordrecht: Kluwer Academic Publishers.

Longino, H. (2002), *The Fate of Knowledge*, Princeton: Princeton University Press.

Massimi, M. (2007). "Saving Unobservable Phenomena". *British Journal for the Philosophy of Science* 58 (2):235-262.

Mayo, D. G. (1996). *Error and the Growth of Experimental Knowledge*. Chicago: The University of Chicago Press.

Mayo, D. G. (2010), "An Ad Hoc Save of a Theory of Adhocness? Exchanges with John Worrall". In Mayo, Deborah G. and Aris Spanos (eds.), *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*, Cambridge: Cambridge University Press: 155–169. doi:10.1017/CBO9780511657528

Mayo, D. G. (2018) *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press.

McAllister, J. W. (1997). "Phenomena and patterns in data sets". *Erkenntnis, 47*, 217–228.

McAllister, J. W. (2011) "What do patterns in empirical data tell us about the structure of the world?" *Synthese*, 182: 73–87.

Moulines, C.U. (1986) "The Ways of Holism", *Noûs, 20*, 313-330.

Newton-Smith, William H. (2001) "Underdetermination of Theory by Data", in W. H. Newton-Smith, ed., *A Companion to the Philosophy of Science*. Blackwell: 532-36.

Norton, J. D. (2008). "Must Evidence Underdetermine Theory?" In *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited*, ed. M. Carrier, D. Howard, and J. Kourany, Pittsburgh: University of Pittsburgh Press: 17-44.

Okasha, S. (2002). "Underdetermination, holism and the theory/data distinction". *The Philosophical Quarterly, 52,* 303–319.

Pickering, A. (1984) "Against Putting the Phenomena First: the Discovery of the Weak Neutral Current", *Studies in History and Philosophy of Science* Part A 15 (2):85 (1984).

Pickering, A. (1995). *The Mangle of Practice,* Chicago: University of Chicago Press.

Quine, W. V. O. (1951/1953) "Two Dogmas of Empiricism", in *From a Logical Point of View*, Cambridge, Mass.: Harvard University Press pp. 20–46.

Quine, W. V. O & Ullian, J. S. (1970/1978). *The Web of Belief*. New York: McGraw-Hill Inc.

Quine, W. V. O. (1975), "Empirically Equivalent Systems of the World", *Erkenntnis*, 9: 313–28.

Quine, W. V. O. (1990/2003). *Pursuit of Truth*. Cambridge, MA: Harvard University Press.

Soler, L. (2008) "The Incommensurability of Experimental Practices: An Incommensurability of What? An Incommensurability of a Third Type?". In Léna Soler, Howard Sankey, Paul Hoyningen-Huene (eds.) *Rethinking Scientific Change and Theory Comparison: Stability, Ruptures, Incommensurabilities?*, Boston Studies in the Philosophy of Science, volume 255, Boston: Springer, 299-339.

Stanford, Kyle, (2021) "Underdetermination of Scientific Theory", *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/win2021/entries/scientific-underdetermination/>.

Suppes, P. 1962. "Models of Data." In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, ed. E. Nagel, P. Suppes, and A. Tarski, 252–61. Stanford, CA: Stanford University Press.

Turnbull M. G. (2017) "Underdetermination in science: What it is and why we should care", *Philosophy Compass*, 2018;13:e12475.https://doi.org/10.1111/phc3.12475TURNBULL11 of 11

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

Woodward, James (2000). "Data, phenomena, and reliability". *Philosophy of Science* 67 (3):179.

Woodward, J. (2010) "Phenomena, Signal, and Noise", *Philosophy of Science*, Vol. 77, No. 5 (December 2010), pp. 792-803.

Tables

The standard approach to underdetermination

<table>
<tr><th rowspan="1">HOLISTIC THEORETICAL UNDERDETERMINATION</th><th colspan="2">CONTRASTIVE THEORETICAL UNDERDETERMINATION</th></tr>
<tr><td rowspan="3">The same set of observations may or may not support or refute the theoretical hypothesis, depending on the choice of background assumptions.</td><td colspan="2">The available evidence is never sufficient to determine the truth of one theoretical hypothesis over another.</td></tr>
<tr><td>Empirical equivalence</td><td>Confirmatory equality</td></tr>
<tr><td>Alternative theories that are empirically equivalent.</td><td>Alternative theories that are equally well confirmed but not empirically equivalent.</td></tr>
</table>

Table 1: Stanford's (2021) taxonomy of theoretical underdetermination.

Parallelism between theoretical and empirical underdetermination

THEORETICAL UNDERDETERMINATION	**EMPIRICAL UNDERDETERMINATION**
• **Holistic version:**	• **Holistic version:**

<table>
<tr><td colspan="2">The same theoretical hypothesis may or may not be supported/refuted by the same empirical phenomenon.</td><td colspan="2">The same empirical phenomenon may or may not be supported/refuted by the same data.</td></tr>
<tr><td colspan="2">• Contrastive version:
Available evidence is not enough to determine the truth of a certain theoretical hypothesis over another.</td><td colspan="2">• Contrastive version:
Available data are not enough to determine the truth about the occurrence of one phenomenon or another</td></tr>
<tr><td>Empirical equivalence</td><td>Confirmatory equality</td><td>Empirical equivalence</td><td>Confirmatory equality</td></tr>
<tr><td>The same empirical phenomena can be interpreted as caused by alternative theoretically postulated phenomena.</td><td>Different empirical phenomena support different inferences to alternative theoretically postulated phenomena.</td><td>The same data can alternatively be interpreted as caused by alternative empirical phenomena.</td><td>Different bodies of data support different inferences to alternative empirical phenomena.</td></tr>
</table>

Table 2: taxonomy of underdetermination.