Highlights

How uncertainty and underdetermination allow measurement to produce useful results

Pierre-Hugues Beauchemin, Kent W. Staley

- We present a philosophically pragmatist conception of measurement as inquiry.
- The problem of usefulness is a central epistemological problem for measurement.
- Uncertainty and underdetermination are crucial to solving the problem of usefulness.
- We reveal the positive epistemic value of underdetermination and uncertainty.

How uncertainty and underdetermination allow measurement to produce useful results

Pierre-Hugues Beauchemin[a], Kent W. Staley[b,*]

[a]*Dept. of Physics & Astronomy, Tufts University, 574 Boston Ave., Medford, MA 02155, U.S.A.*
[b]*Dept. of Philosophy, Saint Louis University, 3800 Lindell Blvd., St. Louis, MO 63108, U.S.A.*

Abstract

Recent philosophical literature on the epistemology of measurement has relegated measurement uncertainty to a secondary issue, concerned with characterizing the quality of a measurement process or its product. To reveal the deeper epistemological significance of uncertainty, we articulate the *problem of usefulness*, which is concerned with the tension between the specificity of the conditions under which particular measurements are performed and the broader range of conditions in which measurement results are intended to be – and are – used. This is simultaneously an epistemological and a practical problem. To articulate the problem and explain its solution we employ a philosophically pragmatist framework that treats measurement as a form of inquiry. Drawing on that framework, we claim that measurement uncertainty is crucial to understanding how in practice investigators solve the problem of usefulness. Explaining exactly how that works, however, yields a surprising result. The contribution of measurement uncertainty to the solution of the problem of usefulness exploits the underdetermination of measurement procedures by the aims of a measurement and the resources available for performing that measurement. *Underdetermination* of measurement, its treatment in terms of the investigation of *uncertainty*, and the relationship of uncertainty to *sensitivity*, are key to enabling investigators to successfully complete measurement inquiries. Our account thus shows how two features of scientific inquiry typically thought of in epistemically negative terms – uncertainty and underdetermination – promote positive objectives in the pursuit of knowledge.

Keywords: uncertainty, underdetermination, sensitivity, pragmatism, inquiry, usefulness, epistemology

*Corresponding author: kent.staley@slu.edu

1. Introduction

In the growing philosophical literature on the epistemology of measurement of the last two decades or so, the trend has been to prioritize accounts of measurement itself, asking such questions as: how is a measurement scale coordinated with a quantity [1, 2]? how does the operationalization of a measurement relate to the quantity being measured [2, 3]? how are measurement procedures validated, and what is the meaning of validity implied by such operations [4, 5]? what makes measurement results objective [6]? do the quantities targeted by measurement procedures have determinate values independently of measurement procedures and their representations [7, 8]? As important as these issues are, *measurement uncertainty* has been treated as a secondary issue, concerned with characterizing the quality of a measurement process or its product.

We argue for a deeper epistemological significance of uncertainty. Our argument rests in part on the articulation of an epistemological problem, which we call the *problem of usefulness*, which is concerned with the tension between the specificity of the conditions under which particular measurements are performed and the broader range of conditions in which measurement results are intended to be – and are – used. The problem of usefulness is simultaneously an epistemological and a practical problem. We claim that measurement uncertainty is crucial to understanding how in practice investigators solve the problem of usefulness.

Explaining exactly how that works, however, yields a surprising result. The contribution of measurement uncertainty to the solution of the problem of usefulness exploits the *underdetermination* of measurement procedures by the aims of a measurement and the resources available for performing that measurement. Insofar as uncertainty estimation reflects such underdetermination at the level of measurement procedures, it enables measurement results to be expressed in a manner that supports investigators managing the underdetermination of theory by the evidence produced by measurement, in an epistemically responsible manner. Our account thus shows how two features of scientific inquiry typically thought of in epistemically negative terms – uncertainty and underdetermination – promote positive objectives in the pursuit of knowledge.

Crucial to our solution to the problem of usefulness is a view of measurement as a kind of inquiry, conceptualized in broadly pragmatist terms. By understanding measurement as inquiry, we make explicit the epistemological significance of measurement objectives, promoting them from playing implicit roles as *merely* contextual matters to essential factors in understanding how measurement contributes to our knowledge of the world. As we explain in Sect. 5, the pragmatist orientation of this account provides the connection between the usefulness of measurement results and their ability to contribute to such knowledge. The problem of usefulness is not, therefore, a narrow problem of a particular corner of the scientific enterprise. It is the problem of how measurement produces knowledge.

We outline our conception of measurement as inquiry in Sect. 2 and the problem of usefulness in Sect. 3. Sect. 4 provides contextualization in relation

to the literature on underdetermination in philosophy of science. We elaborate on our pragmatist account of inquiry in Sect. 5, in order to clarify concepts that we employ in our solution to the problem of measurement. That solution is then provided in the next two sections, with Sect. 6 explaining how uncertainty estimation provides a response to the underdetermination of measurement procedures and Sect. 7 showing how uncertainty, through its close relationship with sensitivity, helps scientific inquiry respond to problems of underdetermination regarding theory in an epistemically responsive manner, enabling measurement to produce useful results that thereby contribute to scientific knowledge.

2. Measurement as inquiry

We treat measurement as a kind of inquiry. That is to say, to measure is to conduct an inquiry, and a measurement result constitutes an outcome of such an inquiry. In Sect. 5 we elaborate on our pragmatist account of inquiry in general. Here we restrict ourselves to a brief characterization of measurement as inquiry: The activity of measurement aims to learn what value(s) one may attribute to something treated as evaluable by a community of scientists and subject to normative constraints. The normative constraints on measurement might feature some field-specific variants, but in all cases of the sort that concern us here[1] a successful measurement will rely on some measuring instrument and produce a result that includes a claim about values attributed to a targeted quantity called the measurand. There are epistemic, in addition to pragmatic reasons, for conducting a measurement process leading to such an evaluation of a measurand; something will be done with such a measured value. In particular, we focus on the *epistemic* use of measurement results: the outcome of a measurement inquiry serves as a resource for the pursuit of a further inquiry.

Because it provides them with a resource to be deployed, measurement leads to a change in the inquirers' relation to their environment and their ability to navigate (in a broad pragmatic sense) in that environment.[2] Hence, measurements constitute inquiries in the sense that we detail further in Sect. 5, and lend themselves to an epistemological analysis in terms of that account.

3. The problem of usefulness of measurement

Two prominent features of scientific measurement appear to be in tension with one another:

[1] Broadly speaking, this includes measurement in experimental physics.

[2] For example, a particular experiment might require that a temperature gradient must be constant within a certain degree of tolerance across some volume of space. Measuring the temperature in some sample of locations within that space will facilitate experimenters performing the experiment according to their own protocol (by enabling them to establish a relationship of control over the temperature gradient) and draw appropriate inferences on the basis of the data they collect.

1. In measuring, inquirers aim to achieve a result that enjoys evidential support.
2. The results of measurement are meant to be useful as evidential support in subsequent inquiries.

Achieving (1) is necessary for (2), but in ways that can limit the extent to which (2) is achieved. On the one hand, the content of measurement results and their evidential support depend on details of the specific conditions of their production. Both the content of the result and the nature of the evidence supporting that result are relevant to the question of what use can be made of the result. Yet, measurement results are produced for the purpose of being used in inquiries and other activities that will be conducted in conditions distinct from those in which they were produced. This tension between dependence upon the conditions of production and the aspiration for usefulness beyond those conditions gives rise to what we call "the problem of usefulness": How do specific and concrete measurement procedures executed in one context produce results that can be used in the conduct of scientific inquiry in a broader range of contexts?

Like the Roman god Janus, the problem of usefulness faces both forward and backward, in a way that can be expressed by two further questions:

A. How do investigators warrant that they appropriately used evidence from previous or ancillary inquiries to inform their choices about how to produce measurement results?

B. How do investigators warrant that the results they produce can be used as evidence in the context of distinct inquiries with their own varied aims?

We propose that understanding the solution to the problem of usefulness of measurement involves a new appreciation and re-evaluation of a more familiar problem in the philosophy of science: the underdetermination of theory by evidence.

4. Problems of underdetermination

Generically, one may think of underdetermination in terms of a relationship among resources and objectives.[3] The resources may include data, observations, claims about phenomena or regularities, or anything that may be regarded as an input to some procedure for making a choice amongst alternatives. The objectives concern the nature of the choice to be made on the basis of those

[3]Our formulation is more general than most references to underdetermination in the philosophy of science literature, which typically focus on the case where the choice to be made concerns theoretical claims and the resource that is considered relevant for that choice is described as 'evidence', 'observations', or 'data.' Our argument invokes instances of the underdetermination relation operating at mulitiple stages in the scientific process and involving multiple kinds of choices.

resources. The choice procedure may be thought of as structured by rules, constraints, or guidelines. Underdetermination obtains when the resources, along with the rules, constraints, or guidelines, are not sufficient to single out a unique choice among the alternatives.

In philosophy of science, claims about underdetermination have often targeted those situations in which some body of data, along with some conception of rules derivable from a deductive logic, fails to determine a unique generalization from those data. Statements of both the "problem of induction" and "Duhem's thesis" [9] often exemplify this tendency [10, 11, 12]. Other versions of underdetermination problems are also prominent, such as Thomas Kuhn's argument that the shared values that scientists rely on when choosing amongst competing theories (e.g. simplicity, consistency, accuracy, fruitfulness, and scope) are insufficient to single out a unique preference [13].

In whatever form and for whatever purpose underdetermination has been invoked in philosophy of science, the claims about it have had a negative tone: underdetermination is a *problem to be solved* or a basis for *skepticism, pessimism, or reluctance to commit*. We cannot rely on evidence to decide what to believe because the available evidence can be invoked to support alternatives to any choice one might make [12]. Multiple theories, including some not yet conceived, are supported by the same evidence. Underdetermination also poses a methodological problem for the practice of science: in absence of a recipe to tell us what to do when evidence appears to count against a theory (although "Duhem's thesis" counsels us not to take this appearance at face value), what should we do to reconcile the evidence with our theoretical commitments? This is often combined with a form of *holism* posing a problem for the specificity of evidence: the evidence can only confirm or disconfirm entire bodies or networks of theories or beliefs; we cannot use it to render judgments on particular claims.

Philosophers have responded to underdetermination in various ways. Some have used the conjunction of underdetermination and holism to revise or even reject the rationality of science (arguably Thomas Kuhn [14] and Paul Feyerabend [15]). Social constructivists have relied on underdetermination by evidence in arguing that disagreements over theories get resolved not by appeals to independently warranted evidence but pressures exerted through social mechanisms ([16, 17, 18]). For W.v.O. Quine [19, 12], the diagnostic is more nuanced: the rationality of science resides in its continuous attempt to bring cohesion to one's beliefs as new sensory inputs prompt revision. But any rules guiding such revisions are themselves part of the beliefs that may revised, so that any considerations of rationality in the process will be pragmatic.

In an important critique, Larry Laudan [20] seeks to deflect the more skeptical and radical epistemological conclusions drawn from underdetermination, recasting Duhem's thesis as a logical truism that could only lead to skepticism over the possibility of a normative theory of scientific rationality if one thought that such a theory could only draw upon the resources of deductive logic, and not on other methodological norms. Although Laudan takes the possibility of a richer methodology (both naturalistic and normative) to offer a sharp alternative and rebuttal to Quine's proposal, he does not address directly

the possibility that the grounds for methodological norms that could help resolve underdetermination problems might themselves rest on just the sort of pragmatic considerations that Quine's account invokes (see [12]).

Whatever one makes of these responses, one will not find among them a positive role for underdetermination in the warranting of scientific knowledge. We propose just such a perspective on underdetermination here. Underdetermination in measurement provides the flexibility needed to make evidence epistemically useful. It does this by allowing for an adaptive response to the tension between the two faces of the problem of measurement. The key to our approach is to recognize two levels at which underdetermination can and does arise: at the level of the relation between theories and the evidence invoked to support or undermine them, and at the level of the relation between measurement procedures and the aims, resources, and methods available to investigators performing measurements. Our account thus focuses on underdetermination from the point of view of the consistency of the evidence with many scientific claims, as well as the underdetermination of measurement procedures. We will address this problem using the language of uncertainties and sensitivities.

5. A pragmatist framing: inquiry

We approach experimental inquiries from a pragmatist perspective in which knowledge is a product of successfully executed processes of inquiry by a community of inquirers. The pragmatic significance of such knowledge rests on its forward-looking stability and suitability for use as a resource in future episodes of inquiry. In this perspective, the problem of usefulness is the problem of how inquiry produces knowledge, i..e how an inquiry produces results that become resources to be used in future inquiries. The problem of underdetermination concerns the sufficiency of the resources and tasks of inquiry for bringing inquiry to a determinate conclusion. To address these problems, understood as problems of inquiry that arise within experimental scientific practice, requires a general epistemological modeling of the process of inquiry, suitable to account for scientific experimental practices.

To this end, we analyze experimental inquiries as a sequence of tasks, i.e. as actions carried out in order to accomplish some aims or objectives. Tasks can be analyzed at different levels of precision and scrutiny and may include such varied actions as analyzing a data sample, manipulating an instrument, writing a report, estimating the signal gain of a photomultiplier tube, or estimating the systematic uncertainty on an estimate of the red-shift of a specific galaxy. There is not a unique way to define, at any level, the tasks performed in an inquiry process. What is essential is that the analysis is sufficient to account for the practice under study and enable assessment of the epistemological claims obtained from it.

Performing a task requires using resources. Resources include data, hardware, background knowledge, know-how, etc. Resources also can be epistemologically analyzed at different levels of resolution, offering different perspectives for making sense of scientific practices.

The main epistemic outcome of an experimental inquiry consists in producing evidential claims by performing tasks with a selective set of resources. We refer to this aspect of an inquiry as the *use mode of the inquiry*. For such an outcome to take the form of a "warrantably assertible" judgment, in the sense of Dewey [21], that claim, along with the entire process that leads to it, needs to be critically assessed. This is the goal of the *critical mode of inquiry*. This mode of inquiry is also accomplished through tasks using resources aiming at some objectives. Tasks, resources, and aims are all critically evaluated in relation to one another because criticism can result in revision of tasks, resources, and/or aims of both the use mode and the critical mode of inquiry. The use and critical modes are therefore not separate activities but are entangled, providing different ways of understanding inquiries and feeding back into one another. This model allows mapping context-specific elements of an experiment to big-picture aspects of scientific inquiries, linking proximate to distant aims.[4]

6. Underdetermination and usefulness: A tension

Underdetermination with respect to the relation between theories and the evidence relevant to them warrants us to be cautious in our claims about what our best theories provide and how they are accepted. Maybe underappreciated is the idea that being cautious about what our evidence supports rests on an underdetermination regarding how evidence is produced, thus warranting caution about what we claim the evidence is.[5] What is targeted when evidence is being produced depends on how experimental inquiries are conducted to yield the evidence, what resources are used, and which theories investigators aim to scrutinize. Paraphrasing the problem of underdetermination as formulated by Kyle Stanford [23] but applied to evidence itself, the question is: given that we are warranted to believe that our best theories, instruments, and experimental techniques used to produce evidential claims can, and eventually will, be supplanted, how and why are we nevertheless committed to accept such evidence as a means to support theories?

The solution we propose to this problem is that scientific practices integrate the variability of the conditions of production of evidence within the evidence itself when performing an inquiry in its critical mode. This variability corresponds to an underdetermination regarding the conduct of a measurement procedure: Supposing the aims of a measurement to be given, and supposing the resources available to those performing the measurement and the tasks they are capable

[4] We take a pluralist stance regarding models of inquiry. Our own model is meant to provide a framework for elucidating how scientific practices of inquiry contribute to the production of knowledge. Other accounts might carve up the conceptual space differently.

[5] By situating underdetermination within an analysis that includes measurement procedures, we are restoring an orientation found in the work of Duhem himself. As Karen Merikangas Darling explains [22], Duhem's own argument regarding underdetermination is grounded in the practice of science, particularly an analysis of measurement and scientific language.

of executing to also be given, the exact procedure to be used in performing the measurement would remain underdetermined.

Suppose, for example, that one wants to use a caliper to measure an object (the "workpiece") along a particular dimension, such as a diameter. The calibration function that allows one to infer a diameter from the indication on the caliper may include several parameters requiring additional data as inputs to the measurement, such as the temperature of the workpiece, the temperature of the caliper's scale, the roughness of the contact between caliper legs and workpiece, the Abbe-error or "wiggle room" between the caliper legs, among other possible factors. (See Tal [24] for an insightful discussion of this type of example.) Given that including any one of these factors in the calibration function demands additional resources be spent on obtaining the values of these factors, the measurer has a choice to make regarding which factors to include. Their choice will make a difference to the definition of the calibration function, which will make a difference to how one arrives at a result, which may make a difference to the measurement result itself.

Moreover, when one acknowledges that aims, resources, and tasks are in fact never simply given but are themselves to some degree contingent upon decisions made with respect to measurement procedure, it becomes unavoidable that for any given measurement, in any given context, choices regarding measurement procedure can be made in multiple ways. The question of how to perform a measurement constitutes a problem of coordinating multiple choices.

As argued in [25], the methodology of evaluating uncertainty provides a way of securing the evidence that measurement produces by varying the potential tasks and resources chosen in the process of producing a measurement result and analyzing the consequences of such variation so as to arrive at a measurement result in the form of an interval that accounts for those consequences. As detailed through an example from experimental high energy physics in [26], this process involves careful consideration of a wide variety of different choices that could have been made to obtain the measurement results, drawing upon different theoretical assumptions, and considerations of instrumental performance, for example. Estimating uncertainty allows investigators to characterize, qualitatively or quantitatively, the variability of experimental evidence resulting from its conditions of production. The objective of this process is two-fold:

1. Warranting that the investigators appropriately used evidence from previous or ancillary inquiries to inform their choices about how to produce measurement results;
2. Allowing the measurement results to "travel" from the conditions of their production to the conditions of their potential use, i.e. allowing future users that might have made different choices, relied on different assumptions, used different instruments, applied different corrections, and so on to legitimately use the evidence as resources for their own new scientific inquiries.

These objectives address, respectively, the two facets (A and B) of the problem of usefulness as described in Sect. 3.

This is the first step in converting underdetermination from a problem into an asset: by assessing the range of values encompassed by an uncertainty evaluation accounting for the differences that could arise from different, underdetermined choices, inquirers conduct a securing process warranting their evidential claims, while simultaneously making these claims useful to a wide range of investigators who might have made other choices, including some that have not even been explicitly formulated. Guided by norms governing the responsible reporting of uncertainty,[6] investigators produce evidence in a manner that acknowledges how it is underdetermined by its conditions of production. Not all measurement contexts demand explicit treatment of uncertainty. In some contexts, it might even be unimportant; for example, the point of performing a measurement might be to show that one is capable of performing a particular procedure, and the result itself is not particularly important. However, by choosing to adhere to norms that encourage investigating uncertainty and incorporating an estimate of it into a measurement result, investigators are able to produce results that can be used as evidence with the purpose of producing knowledge, i.e. for rendering the evidence forward-looking.

This however results in a tradeoff: the consequence of producing evidence that is adaptable to various experimental contexts also makes it compatible with different theoretical claims, making it less capable of differentiating among them. This loss of discriminatory power has the character of the kinds of limitation on empirical import emphasized by more traditional arguments of underdetermination.[7] Underdetermination therefore seems to be in tension with the usefulness of the evidence: to legitimately use evidence as support for scientific claims requires limiting its usefulness by making it amenable to various alternative claims.

At first sight, our pragmatist framework does not seem to offer a complete solution to the underdetermination problem, but rather offers an epistemological explanation for the origin and acuteness of the problem: for evidence to be useful for future scientific inquiries, thus constituting scientific forward-looking knowledge, it has to be conducive to an underdetermination of what it could be used for. How could there be a positive epistemic perspective stemming from such a commingling of the usefulness and the underdetermination problems? To answer this, we need to look at details of practice, informed by our pragmatist model of scientific inquiry.

[6] Practitioners promulgate such norms through various mechanisms, including through institutional devices, such as the *Guide to the expression of uncertainty in measurement* (GUM) published by the Joint Committee for Guides in Metrology of the *Bureau International des Poids et Mesures* (BIPM) [27].

[7] This tradeoff mirrors that noted by Staley between the *security* and *strength* of an evidence claim [25]

7. Useful evidence: Uncertainty and sensitivity

The epistemic value of a measurement comes from the fact that measuring a quantity X aims at producing something useful for other inquiries, such as applying a theory with improved knowledge of the value of its parameters, enabling the construction of a model of some system, or testing among competing hypotheses. For any measurement, there is a set of objectives driving the design, the performance, and the evaluation of the success of that inquiry. Through the critical mode of an inquiry, results of a measurement will therefore be evaluated with respect to their proximate and distant aims. The underdetermination in the usability of evidence can only be meaningfully assessed with respect to the objectives of the inquiry producing the evidence or aiming at using the evidence; just what is underdetermined in a given case, and what follows from such underdetermination, is a matter of context [28]. What prevents underdetermination from posing a general problem for the epistemic value of evidence from scientific inquiries is the practices scientists rely on for managing the specifc manifestations of underdetermination that they encounter [29].

In practice, scientists are not prevented from putting the results of inquiry to evidential use on the basis of a logical argument about the possibility of supporting multiple theory claims by the same evidence. Nor do they limit themselves to resorting to additional criteria of theory choice such as theoretical virtues to resolve problems of underdetermination. A crucial practice in the toolbox of scientists for responding to underdetermination is to quantify the *sensitivity* of different scientific claims to the measurement objectives of the inquiries that produce measurement results. Sensitivity relates explicitly to measurement objectives. A measurement result is useful for learning about possibilities in a given domain to the extent that it is sensitive to the differences among those possibilities. In one kind of case, these differences could simply be different possible values of the measurand itself. In that case, the sensitivity would be assessed directly relative to a target measurement uncertainty [30, p. 27]. For example, one might succeed in a measurement of a very small quantity by getting a result that is incompatible with zero, given the uncertainty. In another kind of case, we might consider sensitivity to be concerned with a proposed use of the measurement result. For example, two competing hypotheses might yield different predictions regarding possible values for the measurand. We could then ask whether the measurement result is sensitive to that difference. (These two kinds of cases come together in the case where investigators plan at the outset to test two hypotheses and use the differences in their predictions to set the target measurement uncertainty.)

In this section, to complete our account of the positive value of underdetermination in measurement, we discuss (1) how, within an inquiry aimed at producing a measurement result, sensitivity is assessed on the basis of subinquiries directed at assessing uncertainty, and (2) the ways in which sensitivity considerations affect the aims and conduct of measurement inquiries.

To understand what sensitivity estimates are, recall that the scientific claims made from an experimental measurement include the outcome of an uncertainty

estimate, quoted as an interval of values constitutive of the evidence to be used in future inquiries. This happens through statements of the form: "the measurement value attributed to the measurand X is $x \pm \Delta x$".[8] A sensitivity limitation happens, for example, when the resulting uncertainty is such that two theories, T_1 and T_2, could both be considered as being supported by the same experimental evidence, i.e. that the difference between T_1 and T_2 with respect to their predictions regarding X is smaller than $2\Delta x$.[9] Two failure scenarios, with distinct objectives, might ensue: In the first, the objective of the inquiry that produced the measurement is to distinguish between these theories, using the result of the measurement of X. In this case, the conduct of the measurement inquiry will either need to be adjusted to be able to discriminate between the two theories or else investigators will need to adopt a different objective. In the second, deciding between the two theories is the objective of some other inquiry beside the one producing the measurement result. In such a case, relying on such an insensitive measurement result would be a poor choice of a resource to use in the conduct of that inquiry. Investigators would need either to adopt a different objective or to somehow gain access to a more sensitive measurement result.

Typical scientific inquiries would not merely aim at determining if T_1 and T_2 are consistent with a measurement result $x \pm \Delta x$, but would instead aim at evaluating the level of support or severity of testing that derives from confronting each theory with the evidence. This can be done in multiple ways, using for example different statistical metrics such as the performance of a χ^2 test, or performing a profile likelihood fit. Each approach relies on a comparison of theories, in terms of their experimental consequences, with the data, in terms of a measurement result drawn from the data that includes the uncertainty; the comparison thus takes account of the uncertainty, which has a bearing on the strength of the evidence for or against the theories of interest. In the absence of identifiable empirical differences between two theories, an inquiry would not aim to distinguish these theories on the basis of evidence.

It could well happen that a measurement fails to distinguish between two theoretical claims although it aims to do so, because it is insensitive to the distinction between those claims. In such a case the underdetermination of the theories in question would not have been resolved. The epistemic value of the inquiry to which that measurement contributed would not, however, necessarily be dismissed. The reciprocal feedback mechanism of an inquiry in its different modes, evaluated with respect to the epistemic objectives of the inquiry, would specify the conditions in which such underdetermination could be relieved. This

[8] The uncertainty interval need not be symmetric in this way, but we assume this for simplicity of expression.

[9] Predictions from theories have uncertainties of their own, adding a nuance that we set aside for the purposes of this discussion. Our argument does not in any way depend on omitting this feature, which is typically dealt with by taking into account uncertainties of both the measurement result and the theoretical prediction when assessing relations of compatibility or support.

could take several forms, for example: one might re-estimate the uncertainties in a way that would render the evidence sensitive to the theory difference targeted by the measurement (perhaps a new method for estimating uncertainty has been found to make more efficient use of data in estimating backgrounds than the method originally used); one might redesign the measurement process to increase the sensitivity to the effect targeted by the measurement (perhaps a redesignation of data selection criteria provides sharper discrimination between signal and background); one might change objectives, so the evidence serves some new epistemic function (perhaps the measurement result may not allow investigators to discriminate the two theories originally, but it does allow them to demonstrate the feasibility of a novel measurement procedure). In all these cases, underdetermination with respect to the questions that the evidence can be used to answer would be relieved, enabling knowledge production of some sort. If none of these solutions were reachable within a given inquiry, the quantification of the sensitivity of the measurement results to the targeted objectives, such as discriminating between T_1 and T_2, would indicate how much improvement in sensitivity a new inquiry would require to achieve the desired epistemic objective.

Crucially for such guidance in future measurement inquiries, uncertainty estimates are themselves the outcomes of inquiries, from which one may learn what aspects of the measurement process contribute most to the uncertainty. If "statistical" uncertainty is dominant, simply collecting more data may suffice. But typically and more interestingly, learning from such inquires about the structure of the uncertainty on a measurement, and the relative size of the various contributions to it, allows future measurements to direct attention to those uncertainty contributions the reduction of which will contribute the most to improved sensitivity and hence the advancement of the epistemic objectives of the measurement [31].

As a consequence, regardless of the limitation in the potential use of a measurement result in future inquiry resulting from the size of the uncertainty, there will almost always be a set of inquiries the objectives of which would be attained by using such a result (or a new result from some modification of the measurement process) as evidence. The underdetermination on the use of evidence will not therefore generally be prohibitive of knowledge production. That is not surprising given the very large number of inquiries in science that are deemed successful by the members of the relevant scientific communities.

We have argued (in Sect. 6) that underdetermination as it arises in a measurement procedure provides a rationale for performing inquiries into the conditions of the measurement performance and the possible variability of outcomes related to those conditions. Expressing measurement results in the form of uncertainties enables those results to become useful by taking account of the differences that could arise from different measurement choices. Incorporating such uncertainty estimates secures the evidence, leading to warranted claims about measurement results, while also widening the range of contexts in which measurement results can be used. In this Section, we have argued that such uncertainty estimates relate also to the sensitivity of measurement

procedures to measurement objectives, and that sensitivity considerations enable measurement inquiries to respond to obstacles in order to produce useful results. Specifically, in cases where measurement objectives are at risk of not being met, judging the sensitivity of the measurement based on its uncertainty can provide the basis for adjustments to tasks, resources, or objectives of the measurement inquiry so that underdetermination obstacles can be overcome and measurement activities may be executed – or at least attempted – that can meet their objectives. Eventually, inquirers will generally adopt some objectives for which sensitivity will not be defeated by the uncertainty stemming from underdetermination. In such cases, underdetermination will have played a useful role in producing measurement results, allowing for knowledge production.

8. Conclusion

We have used a pragmatist approach to the production of knowledge in experimental science to demonstrate how the uncertainty estimates of measurements, corresponding to the underdetermination of empirical evidence by its conditions of production, have strong positive epistemological import.

Uncertainty estimates legitimate the evidential claims to be made by securing evidence while warranting their epistemic transportability. They do this by accounting for the contingency of experimenters' choices about specific tasks and resources in the performance of measurement inquiries, thus enabling such results to "travel" [32] to new experimental inquiries making the evidence usable in different contexts. Underdetermination, which is the source of this uncertainty, is therefore positive insofar as it secures evidence and allows using measurement results in different contexts. The uses themselves are inevitably limited by the uncertainty, given that greater uncertainty entails less sensitivity, but attending to this relationship between uncertainty and sensitivity is also the key to maintaining the epistemic value of measurement results in the face of underdetermination of theory by evidence. As a consequence, the limitations arising from underdetermination and uncertainty are merely an aspect of the practical nature of measurement and not a basis for epistemological skepticism or pessimism. Indeed, the uncertainty of measurement results enables them to be used to confront theories empirically in a manner that takes into account underdetermination at two levels: the underdetermination of measurement results by the conditions of their production and the underdetermination of theory by evidence. The former underdetermination is directly reflected in the size of the uncertainty included in the measurement result, which also then provides a quantitative basis for the epistemic caution necessitated by the latter underdetermination. Underdetermination allows investigators to secure scientific evidence and transport measurement results to new contexts, without preventing inquiries from meeting the objective of producing new knowledge.

In making this argument, we have taken a pragmatist philosophical approach. We have shown how placing evidence in the context of the epistemic objectives of inquiry allows us to understand how investigators, confronted with un-

derdetermination, produce knowledge, understood as retrospectively warranted and prospectively useful resources, to be used in other new inquiries.

This work thus offers a positive perspective on uncertainty and underdetermination as essential components in legitimating the role evidence plays in the production of scientific knowledge and addressing the problem of the usefulness of measurement in science.

References

[1] B. van Fraassen, Scientific Representation: Paradoxes of Perspective, Oxford University Press, Oxford, 2008.

[2] H. Chang, Inventing Temperature: Measurement and Scientific Progress, Oxford Studies in Phiosophy of Science, Oxford University Press, New York, 2007.

[3] H. Chang, Operationalism: old lessons and new challenges, in: N. Mößner, A. Nordmann (Eds.), Reasoning in Measurement, Routledge, New York, 2017, pp. 25–38.

[4] D. Borsboom, G. Mellenbergh, J. van Heerden, The concept of validity, Psychological Review 111 (4) (2004) 1061–1071.

[5] U. Feest, Construct validity in psychological tests – the case of implicit social cognition, European Journal for Philosophy of Science 10 (2020) 4.

[6] E. Tal, A model-based epistemology of measurement, in: N. Mößner, A. Nordmann (Eds.), Reasoning in Measurement, Routledge, New York, 2017, pp. 233–253.

[7] P. Teller, Measurement accuracy realism, in: I. F. Peschard, B. C. van Fraassen (Eds.), The Experimental Side of Modeling, University of Minnesota Press, Minneapolis, Minnesota, 2018, pp. 273–298.

[8] J. E. Wolff, Coordination and obsolescence: a response on behalf of measurement realism, Synthese 201 (2023) 78.

[9] P. Duhem, The Aim and Structure of Physical Theory, Princeton Science Library, Princeton University Press, Princeton, NJ, 1991.

[10] L. Henderson, The Problem of Induction, in: E. N. Zalta, U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy, Winter 2024 Edition, Metaphysics Research Lab, Stanford University, 2024.

[11] P. Lipton, Induction, in: M. Curd, J. A. Cover (Eds.), Philosophy of Science: The Central Issues, Norton, 1998.

[12] K. Stanford, Underdetermination of Scientific Theory, in: E. N. Zalta, U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy, Summer 2023 Edition, Metaphysics Research Lab, Stanford University, 2023.

[13] T. Kuhn, Objectivity, value judgment, and theory choice, in: The Essential Tension, University of Chicago Press, Chicago, 1977, pp. 320–339.

[14] T. Kuhn, The Structure of Scientific Revolutions, Third Edition, University of Chicago Press, Chicago, 1996.

[15] P. Feyerabend, Against Method, revised Edition, Verso, London, 1988.

[16] H. M. Collins, Changing Order: Replication and Induction in Scientific Practice, Chicago University Press, Chicago, 1992.

[17] A. Pickering, Constructing Quarks: A Sociological History of Particle Physics, Chicago University Press, Chicago, 1984.

[18] S. Shapin, S. Schaffer, Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life, Princeton University Press, Princeton, NJ, 1985.

[19] W. v. O. Quine, Two dogmas of empiricism, Philosophical Review 60 (1951) 20–43.

[20] L. Laudan, Demystifying underdetermination, in: C. W. Savage (Ed.), Scientific Theories, University of Minnesota Press, 1990, pp. 267–97.

[21] J. Dewey, Logic: The Theory of Inquiry, Henry Holt and Company, New York, 1938.

[22] K. M. Darling, The complete Duhemian underdetermination argument: Scientific language and practice, Studies in History and Philosophy of Science 33 (2002) 511–533.

[23] P. K. Stanford, Refusing the devil's bargain: What kind of underdetermination should we take seriously?, Philosophy of Science 68 (2001) S1–S12.

[24] E. Tal, The epistemology of measurement: A model-based account, Ph.D. thesis, University of Toronto, Toronto (2012).

[25] K. W. Staley, Securing the empirical value of measurement results, British Journal for the Philosophy of Science 71 (1) (2020) 87–113. doi:10.1093/bjps/axx036.

[26] P.-H. Beauchemin, Autopsy of measurements with the ATLAS detector at the LHC, Synthese 194 (2017) 275–312. doi:10.1007/s11229-015-0944-5.

[27] Joint Committee for Guides in Metrology Working Group I, Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement, Joint Committee for Guides in Metrology, 2008.
URL `http://www.bipm.org/en/publications/guides/gum.html`

[28] P. D. Magnus, Background theories and total science, Philosophy of Science 72 (2005) 1064–1075.

[29] S. Ritson, K. Staley, How uncertainty can save measurement from circularity and holism, Studies in History and Philosophy of Science 85 (2021) 155–165.

[30] Joint Committee for Guides in Metrology Working Group II, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, Joint Committee for Guides in Metrology, 2012.
URL http://www.bipm.org/en/publications/guides/vim.html

[31] S. Ritson, Creativity and modelling the measurement process of the higgs self-coupling at the lhc and hl-lhc, Synthese 199 (5) (2021) 11887–11911. doi:10.1007/s11229-021-03317-y.
URL https://doi.org/10.1007/s11229-021-03317-y

[32] P. Howlett, M. S. Morgan (Eds.), How Well Do Facts Travel? The Dissemination of Reliable Knowledge, Cambridge University Press, New York, 2011.