Philosophy of Science (2025), **92**, 1500–1510
doi:10.1017/psa.2025.10147

PSA PHILOSOPHY OF SCIENCE ASSOCIATION

PHILOSOPHY OF SCIENCE

CONTRIBUTED PAPER

Data Can Be Underdetermined, Too

Dana Tulodziecki

Department of Philosophy, Purdue University, United States
Email: dtulodzi@purdue.edu

(Received 24 January 2025; revised 12 March 2025; accepted 24 July 2025; first published online 06 October 2025)

Abstract

This paper focuses on a type of underdetermination that has barely received any philosophical attention: underdetermination of data. I show how one particular type of data—RNA sequencing data, arguably one of the most important data types in contemporary biology and medicine—is underdetermined, because RNA sequencing experiments often do not determine a unique data set. Instead, different ways of generating usable data can result in vastly different, and even incompatible, data sets. But, since it is often impossible to adjudicate among these different ways of generating data, 'the data' coming out of such experiments is underdetermined.

1. Introduction

Underdetermination of theories by data is a longstanding problem in philosophy of science (Laudan 1990): since scientific data is always compatible with a number of different and mutually incompatible scientific theories, it can never, on its own, single out a particular scientific theory uniquely. And while almost every aspect of underdetermination—how ubiquitous it is, how to escape it, what kinds there are, how worrisome these different kinds are, etc.—has been extensively debated (Tulodziecki 2007, 2017), the general phenomenon is familiar and well understood. In this paper, I focus on a type of underdetermination that, contrary to the familiar type, has barely received any philosophical attention at all: underdetermination of data itself.[1]

The idea of data as a potentially philosophically interesting notion has a long heritage. Bogen and Woodward already drew attention to the complexities of data production in the 1980s (1988, 309–10), yet most discussions since have focused on data interpretation (Woodward 1989; McAllister 1997, 2011; Bogen 2010). Recent years

[1] The only discussion of which I am aware that touches on underdetermination of data is Wylie (2019), who argues that fossil specimens—potential data—are underdetermined by context. Note also that I will not, at this point, talk about underdetermination of data *by* something, since different kinds of data can be underdetermined by different things (methods, theoretical assumptions, samples, their production, etc.). It will become clear in Section 3 what exactly the underdetermination of RNA-Seq data refers to.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Philosophy of Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

have seen a new surge of discussions involving data, especially in the context of big data and data-centric approaches to science. However, these discussions, though plentiful, focus mostly on the unprecedented amounts of data produced and associated issues, such as data analysis, data models, data collection, data curation, data infrastructures, or data dissemination. And while it has been argued that data is relational and underdetermines evidential value (Leonelli 2015), even here the focus is on how data is used, not on the epistemic status of data itself. The fact that the latter has been discussed so little is perhaps somewhat surprising in view of the fact that it is also well known among philosophers that data is not given but made (Hacking 1992). It is similarly well known that observation—and, by extension, data—is theory-laden (Hanson 1958). The same point has been argued for experiments, science's prime vehicles for producing data (Franklin 2015; Karaca 2013). Lastly, philosophers of experimentation have also long pointed out that data can be artifacts of the instruments or procedures used to generate them (Rasmussen 1993; Feest 2014). The epistemic status of data would therefore seem to be a natural candidate for philosophical discussion; yet, the literature still contains a puzzling gap in this respect.

My goal in this paper is to begin filling this gap by talking about how one particular type of data—RNA sequencing data, arguably one of the most important types of data in contemporary modern biology and medicine—is underdetermined. I will argue that there is no matter of fact about what 'the data' of many modern RNA sequencing experiments is. Just as in traditional underdetermination evidence does not single out a particular theory, in this case modern RNA sequencing experiments often do not determine a particular data set and therefore leave open what 'the data' coming out of such experiments is. Moreover, as I will show, this underdetermination is not epistemically innocent in the sense that, trivially, slightly different methodological choices give rise to slightly different data sets. Instead, as I will argue, what 'the data' of such experiments is depends so heavily on the ways in which experimental reads are made usable, that different ways of generating usable data can result in vastly different and, in the most extreme cases, even incompatible, data sets.

I will proceed as follows: after providing some background and describing modern RNA sequencing technology (section 2), I will explain some different ways in which the experimental reads coming out of such experiments can be made usable and show that these can result in vastly different, and even incompatible or opposing, data sets (section 3). I then go on to explain why we should think of this as a genuine and serious case of data underdetermination (section 4). Next, I discuss some of the consequences of this type of underdetermination for already existing data, and end by highlighting why philosophers should pay more attention to data underdetermination (section 5).

2. Background and RNA-Seq

RNA is a nucleic acid molecule that carries genetic information for making proteins and regulating gene expression. It is involved in many of the most fundamental biological cellular processes and, since it sheds light on how instructions from DNA are interpreted and subsequently used, data from experiments that sequence RNA are of enormous importance in modern biology and medicine. RNA-Seq, short for 'RNA

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

sequencing', is a recent and powerful high-throughput next-generation sequencing technique that is used to sequence and measure gene expression in different cell types, tissues, organisms, or species, from different developmental stages, or under different experimental conditions (such as healthy vs. diseased tissue (Kruse et al. 2019), treated vs. untreated cells (Calhoun et al. 2022), preserved vs. unpreserved tissue (Kruse et al. 2017), and so on). Gene expression levels measure how active a gene is in producing functional, biological outputs, and RNA transcripts—RNA molecules copied from a gene—function as a proxy for measuring this activity, with higher numbers of transcripts indicating higher activity levels and therefore stronger gene expression, and lower numbers the reverse. RNA-Seq allows one to count all the different RNA transcripts in a sample, and therefore makes it possible to measure which genes are expressed, at what levels they are expressed, and how these expression levels change in response to different conditions or treatments, thereby also making it possible to compare gene expression levels of different samples.

RNA-Seq is used in virtually every life science discipline that deals with biological samples (Lonsdale et al. 2013; Schaum et al. 2018; Zhang et al. 2020), from agriculture and environmental science to neuroscience and precision medicine. It has revolutionized work in these areas, because, unlike previous methods such as microarrays, it can sequence an entire transcriptome (the entire set of RNA molecules in a cell or organism at a particular time, during a particular experimental condition) at once. One major limitation of microarrays is that they can only detect previously known and specified sequences, whereas RNA-Seq can detect and identify novel, previously unidentified transcripts (and previously unknown genes), thereby vastly expanding our knowledge of the transcriptome. Moreover, microarrays can only provide relative rankings of gene expression within a sample and are not able to capture gene expression levels in numerical form, and so RNA-Seq made it possible for the first time both to quantify gene expression levels and to quantitatively compare different samples —all without needing to know anything about the genes in the samples in advance.

So, how does RNA-Seq work? Very briefly, RNA-Seq can 'directly' read nucleic sequences of RNA molecules. It first extracts all the RNA from a sample, then converts the RNA into more stable DNA, in the process creating a library that represents all the RNA molecules of the sample. The sequencer then reads the sequence of bases in the DNA, one constituent base at a time, resulting in millions of short sequences ('reads'). These reads provide a snapshot of the sample's entire transcriptome, i.e. of all the RNA molecules present in the sample. Through processes called 'mapping' and 'quantification', it is then determined which RNA molecules (and hence which genes) the reads came from and how much of each molecule was in the sample, with this amount serving as a proxy for gene expression in the form of 'read counts', i.e. the number of reads for each RNA transcript.

3. Making data usable

However, these 'raw' read counts are unusable in their initial form.[2] To be able to compare samples, the read counts first need to be adjusted for variations in

[2] Before the reads are counted, just like for other biological data and techniques, read errors need to be identified and corrected, low-quality reads need to be sorted out, and so on. It is only after these

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

'sequencing depth', the total number of sequencing reads for a particular sample. Sequencing depth typically varies among samples, due to their respective RNA quality and quantity. For example, if sample *A* has twice as many RNAs as sample *B*, the read counts for sample *A* might be roughly twice as high as those for sample *B*, but this would not necessarily indicate doubled biological gene expression levels.[3] Similarly, if the sequencer sequences more of sample *A* than of sample *B* because the sample quality of *A* is higher, sample *A* will receive more read counts without this indicating biologically higher expression levels. To account for this, and so that the read counts can reflect actual gene expression and not merely sample variation, the read counts of all the samples in an experiment first need to be adjusted to a common base (or scale) via a procedure called 'scaling'. It is only after this is done that read counts are even candidates for analysis and interpretation.

There are currently three main types of scaling method for RNA-Seq.[4] The first involves housekeeping genes. These are genes that are thought to be expressed at a constant level across samples within an experiment and that allow other genes to be scaled in proportion to the reference values of the housekeeping genes. In the second method, the total gene expression method (also sometimes called 'the total count normalization method'), the total number of reads in each sample is used as a scaling factor by which each gene's read counts are divided.[5] In contrast to housekeeping genes, which use only specific genes as reference, total count normalization relies on the totality of mapped reads. The third method uses so-called spike-in controls. These are known biological or synthetic sequences of known concentration that are added to ('spiked into') the sample before sequencing, and this known quantity of RNA can then be used as a reference point.

However, it turns out that these different scaling methods can lead to vastly different data sets. One such example comes from RNA-Seq experiments trying to identify changes in gene expression levels during aging in yeast. While scaling with spike-ins led to data that showed that all genes in the yeast genome were upregulated (more strongly expressed) during aging (Hu et al. 2014), not using spike-ins resulted in data that showed that there were gene expression changes in only a few hundred genes, some upregulated and some downregulated, with the expression levels of most genes unchanged (Chen et al. 2016; Lesur and Campbell 2004). Similarly, using total count normalization to scale read counts for an experiment involving cMyc oncogene—a protein-coding gene that can promote cancer development—produced data that showed that overexpression of cMyc activated only a specific number of target genes (Lovén et al. 2012, 2013). However, scaling using spike-in controls produced data that

cleaned-up reads are assigned or mapped to specific genes that the counting begins. Since issues related to this are not the focus of this paper, I won't pursue them here.

[3] This example assumes an exhaustive sequencing of the two samples.

[4] Each of these comes in a number of variations, but since those details don't matter for my purposes, I will ignore this additional complication here.

[5] There are several variations on this method (see, for example, Cole et al. 2019). The terms 'scaling' and 'normalization' are not used consistently in the literature. I use 'scaling' to refer to procedures that are used to provide a common base for different samples and 'normalization' to refer to methods that need to be implemented even within a single sample to make it usable (such as adjusting for gene length, library size, sequencing depth, and so on). The vast majority of RNA-Seq experiments involve more than one sample and hence both scaling and normalization, so I'll often refer to both.

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

showed that cMyc amplified almost the entire genome (Lin et al. 2012; Nie et al. 2012).[6] Thus, as we can see, different scaling methods can produce incompatible data from one and the same RNA-Seq experiment. Moreover, this issue is not just confined to the three main types of scaling methods, but extends to variations of the various methods within a category. For example, there are a variety of different total count normalization methods and here, too, it has been shown that different methods can produce significantly different data (Bullard et al. 2010; Dillies et al. 2013).

4. The underdetermination of RNA-Seq data

So, why should we think of this as a case of data underdetermination? The answer involves the fact that the different scaling methods all rest on different assumptions and that whether and to what extent an experiment meets these assumptions is usually and largely unknown. To give some quick and easy examples, one assumption underlying the housekeeping gene method is that the chosen set of housekeeping genes is, in fact, constantly expressed in the sample. The total gene expression method presupposes that total expression among different experimental conditions is the same, and spike-in controls presuppose that the spike-ins themselves won't be affected by the biological condition under investigation and, in the case of synthetic spike-ins, also that "they have the same technical effects as real genes" (Evans et al. 2018, 781). And while each of these assumptions has been known to be violated sometimes, there are "many situations in which the validity of any assumption is unknown for the given experiment" (790). Further, not just does there "not exist an ... analysis of published data, which evaluates the assumptions," there is also "no clear way to perform such an evaluation" (791).

The data coming out of many RNA-Seq experiments is therefore underdetermined: to choose the most appropriate scaling method (i.e. the method most likely to produce data actually reflecting the sample's biological expression levels), one would need to know to what extent the various assumptions underlying the different scaling methods are met in the experiment at hand, when in many cases this is impossible. In the classical case of underdetermination of theories by evidence, underdetermination obtains when there are two or more incompatible theories among which we can't adjudicate on evidential grounds because they are all compatible with the observable evidence. Here, underdetermination of data obtains due to the fact that the read counts that come out of one and the same sequencing run are compatible with differently scaled, potentially even opposing, data sets among which one can't adjudicate. It is thus, in principle, unclear what the best scaling method is, and given that different methods lead to different data, the data in these cases is genuinely underdetermined. Since differently scaled data sets are all compatible with the original sample, there is no matter of fact about what 'the data' of such an experiment is.

One might think that one obvious step towards resolution would be to simply use several methods at once. But this would help only to a limited extent: if it turned out that the methods agreed, one could be confident in the general data trends; if they disagreed, however, one would know to worry, but not how to resolve this worry,

[6] For further and different examples, see Lovén et al. (2012), Chen et al. (2016), and Evans et al. (2018).

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

since one still wouldn't know which method's underlying assumptions are best instantiated in a given experiment. At any rate, using several different methods is not usually a live option (even in cases in which it may be a theoretical one), because both sequencing and subsequent analysis are expensive, for many researchers prohibitively so. Only a small number of universities can afford their own in-house sequencing facilities and even in those cases the cost is significant because such facilities inevitably rely on commercial (re)sources for equipment, reagents, and support. If local sequencing is not an option—as is the case for the vast majority of researchers—samples are sent out to commercial facilities, often at still greater cost. Note that in both cases even 'merely' bioinformatic work is quite expensive due to labor costs and so using different methods is not usually realistic.

Regardless, the foregoing discussion might lead one to wonder who usually determines which scaling and normalization methods are used. While some researchers might be actively involved in such decisions, more often than not the researchers designing the experiments have little to no expertise in any of the methods and techniques involved in producing the experimental data, and often not even in any of the methods required for the subsequent statistical analyses. Instead, researchers often buy kits and packages that outsource the entire sequencing and analysis process to a sequencing facility.[7] They are provided with a standardized kit to prepare their samples, which are then sent to the sequencing facility, where sequencing is done by a technician before one of the resident bioinformaticians deals with the raw (and still unusable) read counts. This includes scaling and normalization procedures, but often also involves the requisite statistical analyses of the scaled and normalized data. When all this is complete, researchers receive a report compiled by the sequencing facility about the fully processed and possibly even analyzed data. Thus, the researchers who design the experiment are often quite removed from the data generation process itself, and many simply send their samples out for sequencing without giving any thought to the methods and procedures involved, much less their underlying assumptions. But not just are many researchers removed from this process practically speaking, they are also often in a field (cell biology, medical physiology, oncology, plant science, agriculture, etc.) that does not come with background, expertise, or even competency in molecular biology or bioinformatics. And since the methods involved in scaling, normalization, and data analysis are heavily mathematical and statistical, they cannot be used or even understood without significant training. The important consequence of this for our purposes is that for most researchers, not just the methods themselves but also their assumptions remain obscure and inaccessible. While some researchers might go out of their way to reflect on and pick specific methods tailored to their experiment, most of the time the bioinformatician employed by the sequencing facility picks the method that will be used. Sometimes labs have their own guidelines about what methods to use when, sometimes a researcher's institution will have guidelines they want followed, sometimes particular methods are used simply because they became entrenched and are "what has always been done." The key point here is that it is relatively rare for a method to be matched specifically to a particular experiment or sample because its underlying assumptions are thought to be most appropriate for that specific case.

[7] For how this affects data production, see Krohs (2012).

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

And, just like cell biologists don't usually have training in bioinformatics, bioinformaticians don't usually have training in cell biology, or the specific subject matter expertise relevant to the researcher's discipline. The requisite kind of tailoring of method to experiment, however, requires both.

This is perhaps also a good point at which to mention that spike-in controls are the least commonly used scaling method. I already mentioned earlier the cost associated with doing RNA-Seq experiments. Spike-in controls not only increase this cost but also require, ideally, robotics for procedure automation as well as additional expertise—not just technical expertise in 'spiking in', but also theoretical expertise in how to choose an appropriate spike-in sequence for a particular sample and experiment. Some facilities have experience with spike-ins, but many don't have staff with the required expertise on site. In those cases, spike-in controls might add another layer of expertise that needs to be brought in externally, for example through companies specializing in such controls.

5. Consequences for already existing data

While RNA-Seq scaling and normalization problems are not unknown in the scientific literature,[8] they and their consequences have been underappreciated. Part of the reason for this might be the following two tendencies: first, the tendency to think that the potential differences in RNA-Seq data are not significant enough to be genuinely problematic; second, the tendency to think of RNA-Seq data as highly reliable. For example, specifically with reference to next-generation sequencing technologies like RNA-Seq, Leonelli has noted "the reliance on specific technologies for data production as proxy markers for data quality" (2017, 4). The perception of reliability might also be especially strong in the case of RNA-Seq because it is not just highly sophisticated, but also heavily automated and standardized, with many detailed protocols. It also explicitly eliminates many of the problems and biases of previous sequencing techniques that depended more strongly on individual researcher usage (although it should be noted, of course, that plenty of the 'usual' biological and technical biases also occur in RNA-Seq). Moreover, where previous methods generated qualitative or at most semi-quantitative data that was in more obvious need of researcher interpretation, RNA-Seq generates entirely quantitative data, a fact that might further contribute to the notion that RNA-Seq data is more objective and less prone to interpreter bias. Perhaps there is even a tacit assumption that numbers are more objective representations of actual biological states of affairs than are researcher interpretations of qualitative data. Further, whereas microarrays relied on pre-designed probes with known sequences, the fact that RNA-Seq measures gene expression more directly and that it can measure an entire transcriptome—including previously unknown genes and transcripts—might lend a further air of objectivity and reinforce the idea that it offers an unbiased view of what the transcriptome 'really' is.

An important point to note here is that because "the use of technology as proxy for data quality continues to occur among editorial boards, research institutions and

[8] There is also evidence that at least some researchers are uneasy about "their use of high-throughput technologies for data production," and especially about "the level of technical skill required to use those tools, [and] the proficiency with which lab members were operating the technology" (Leonelli 2017, 5).

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

funders, and international research consortia" (Leonelli 2017, 5), researchers are in fact incentivized *not* to deviate from the existing standardized procedures of certain facilities—even if this would serve their experiment—since using that facility and its protocols serves as such a data quality proxy.

What, then, are the consequences of all this for existing data sets? In the luckiest case, the researcher who deposited the data was a person who thought about these issues carefully, documented their thoughts, and then went to make these available, along with the untreated read counts and unprocessed data, as well as the scaled/normalized data. But such cases are the exception, not the norm. As we just saw, RNA-Seq experiments are often performed without a particular rationale for using a particular scaling or normalization method. This means that there are enormous quantities of already existing and publicly available data sets for which it is unclear not just why a particular method was used but also whether they are genuinely underdetermined or whether they, at least theoretically, come with a preferred method. Unfortunately, even in principle, this situation can be addressed only partially. As has often been pointed out, "existing databases have a hard time getting data producers to post and appropriately annotate their own data" (Leonelli 2017, 4). So, even if data is available, it might have been made available in sloppy or haphazard ways. For our purposes, this means that often, even if 'the data' is there, it consists only of the scaled or normalized data, not the original untreated read counts. In those cases, it is impossible to retroactively apply different methods, even if it becomes clear that a different method would have been preferable (and, at any rate, spike-in controls need to be added to samples before sequencing and so it is too late for this method, anyway). Moreover, even in cases in which the untreated read counts are available, the experimental metadata might be insufficient for judging whether certain assumptions underlying the different methods are met or violated for a given experiment, even when—with sufficient documentation—this could have been determined.

On a practical level, this means that there is often no way to tell whether an already existing data set is genuinely underdetermined or not. In fact, since this can often not be resolved, there is meta-level underdetermination of judgment about whether there is lower-level epistemic underdetermination at the data level. It is impossible to tell what sort of situation one is in: is one looking at a data set that is such that a different method would have produced data in conflict with what was originally concluded? If so, was one of the methods preferred and was it used for the original data, or was there no preferred method and the data is experimentally underdetermined? Since such questions cannot be resolved, there is no way of knowing how widespread either of these kinds of underdetermination really are. It is practically underdetermined whether the data is epistemically underdetermined, and nobody knows just how bad things are.

The downstream consequences of this situation are not insignificant, especially in an age in which analysis and use of legacy data are becoming increasingly scientifically important and encouraged. Plenty of existing data sets are used in comparative experiments and in influential review papers and meta-analyses. Without seeking to understate or diminish the enormous successes of RNA-Seq and the very important role RNA-Seq has played in many fields, especially medicine, what the discussion of data underdetermination shows is that there is an urgent need to

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

further reflect on the epistemic status of much of this data. Is most of this underdetermination genuine or in principle resolvable? Regardless of the answer, existing data needs to be probed with a view to ascertaining this, and also with a view to ensuring that no assumptions were violated during the scaling and normalization processes. The specter of data underdetermination also raises a number of further questions it is important to shed light on: are there particular types of experiments, specimens, or biological conditions that are especially prone to genuine underdetermination? If so, what, if anything, do they share? Is there a way of overcoming such underdetermination, and if so, how? But even regardless of the question of whether genuine underdetermination occurs, practical underdetermination of existing data is already sufficient to cast doubt on the reliability of 'the data' and to prompt further analysis of its epistemic status.

It should also prompt philosophers both to rethink the epistemic status of data more generally and to think about how data underdetermination might affect widely used and important philosophical concepts, among them empirical equivalence, classical underdetermination, phenomena, and, not least, the notion of evidence itself.

Acknowledgements. Many thanks to Uljana Feest, Chris Haufe, D. Marshall Porterfield, and Hansjörg Schwertz for helpful conversations about this paper. Special thanks to Colin Kruse for helping me understand RNA-Seq, for our many discussions about scaling, and for commenting on the penultimate draft of this paper.

Declaration of competing interests. None to declare.

Funding information. None to declare.

References

Bogen, Jim. 2010. "Noise in the World." *Philosophy of Science* 77 (5):778–91. https://doi.org/10.1086/656006.

Bogen, Jim and James Woodward. 1988. "Saving the Phenomena." *The Philosophical Review* 97 (3):303–52. https://doi.org/10.2307/2185445.

Bullard, James H., Ellizabeth Purdom, Kasper D. Hansen, and Sandrine Dudoit. 2010. "Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments." *BMC Bioinformatics* 11 (1):94. https://doi.org/10.1186/1471-2105-11-94.

Calhoun, Sara, Bishay Kamel, Tisza A. S. Bell, Colin P. S. Kruse, Robert Riley, Vasanth Singan, Yuliya Kunde, Cheryl D. Gleasner, Mansi Chovati, Laura Sandor, Christopher Daum, Daniel Treen, Benjamin P. Bowen, Katherine B. Louie, Trent R. Northen, Shawn R. Starkenburg, and Igor V. Grigoriev. 2022. "Multi-omics Profiling of the Cold Tolerant *Monoraphidium minutum* 26B-AM in Response to Abiotic Stress." *Algal Research* 66:102794. https://doi.org/10.1016/j.algal.2022.102794.

Chen, Kaifu, Zheng Hu, Zheng Xia, Dongyu Zhao, Wei Li, and Jessica K. Tyler. 2016. "The Overlooked Fact: Fundamental Need for Spike-in Control for Virtually all Genome-wide Analyses." *Molecular and Cellular Biology* 36 (5):662–67. https://doi.org/10.1128/MCB.00970-14.

Cole, Michael B., Davide Risso, Allon Wagner, David DeTomaso, John Ngai, Elizabeth Purdom, Sandrine Dudoit, and Nir Yosef. 2019. "Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-Seq." *Cell Systems* 8 (4):315–28. https://doi.org/10.1016/j.cels.2019.03.010.

Dillies, Marie-Agnès, Andrea Rau, Julie Aubert, Christelle Hennequet-Antier, Marine Jeanmougin, Nicolas Servant, Céline Keime, Guillemette Marot, David Castel, Jordi Estelle, Gregory Guernec, Bernd Jagla, Luc Journeau, Denis Laloë, Caroline Le Gall, Brigitte Schaëffer, Stéphane Le Crom, Mickaël Guedj, and Florence Jaffrézic. 2013. "A Comprehensive Evaluation of Normalization Methods for Illumina High-Throughput RNA Sequencing Data Analysis." *Briefings in Bioinformatics* 14 (6):671–83. https://doi.org/10.1093/bib/bbs046.

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

Evans, Ciaran, Johanna Hardin, and Daniel M. Stoebel. 2018. "Selecting Between-Sample RNA-Seq Normalization Methods from the Perspective of their Assumptions." *Briefings in Bioinformatics* 19(5), 776–92. https://doi.org/10.1093/bib/bbx008.

Feest, Uljana. 2014. "Phenomenal Experiences, First-Person Methods, and the Artificiality of Experimental Data." *Philosophy of Science* 81 (5):927–39. https://doi.org/10.1086/677689.

Franklin, Allan. 2015. "The Theory-Ladenness of Experiment." *Journal for General Philosophy of Science* 46:155–66. https://doi.org/10.1007/s10838-015-9285-9.

Hacking, Ian. 1992. "The Self-Vindication of the Laboratory Sciences." In *Science as Practice and Culture*, edited by Andrew Pickering, 29–64. Chicago, IL: University of Chicago Press.

Hanson, Norwood R. 1958. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press.

Hu, Zheng, Kaifu Chen, Zheng Xia, Myriah Chavez, Sangita Pal, Ja-Hwan Seol, Chin-Chuan Chen, Wei Li, and Jessica K. Tyler. 2014. "Nucleosome Loss Leads to Global Transcriptional Up-Regulation and Genomic Instability During Yeast Aging." *Genes and Development* 28 (4):396–408. https://doi.org/10.1101/gad.233221.113.

Karaca, Koray. 2013. "The Strong and Weak Senses of Theory-Ladenness of Experimentation: Theory-Driven versus Exploratory Experiments in the History of High-Energy Particle Physics." *Science in Context* 26 (1):93–136. https://doi.org/10.1017/S0269889712000300.

Krohs, Ulrich. 2012. "Convenience Experimentation." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1):52–57. https://doi.org/10.1016/j.shpsc.2011.10.005.

Kruse, Colin P. S., Proma Basu, Darron R. Luesse, and Sarah E. Wyatt. 2017. "Transcriptome and Proteome Responses in RNAlater Preserved Tissue of *Arabidopsis thaliana*." *PloS One* 12 (4):e0175943. https://doi.org/10.1371/journal.pone.0175943.

Kruse, Colin P. S., David A. Cottrill, and John J. Kopchick. 2019. "Could Calgranulins and Advanced Glycated End Products Potentiate Acromegaly Pathophysiology?" *Growth Hormone & IGF Research* 46–47:1–4. https://doi.org/10.1016/j.ghir.2019.04.002.

Laudan, Larry. 1990. "Demystifying Underdetermination." In *Scientific Theories*, edited by C. Wade Savage,267–97. Minneapolis, MN: University of Minnesota Press.

Leonelli, Sabina. 2015. "What Counts as Scientific Data? A Relational Framework." *Philosophy of Science* 82 (5):810–21. https://doi.org/10.1086/684083.

Leonelli, Sabina. 2017. "Global Data Quality Assessment and the Situated Nature of 'Best' Research Practices in Biology." *Data Science Journal* 16:32. https://doi.org/10.5334/dsj-2017-032

Lesur, Isabelle and Judith L. Campbell. 2004. "The Transcriptome of Prematurely Aging Yeast Cells is Similar to that of Telomerase-Deficient Cells." *Molecular Biology of the Cell* 15 (3):1297–312. https://doi.org/10.1091/mbc.e03-10-0742.

Lin, Charles Y., Jakob Lovén, Peter B. Rahl, Ronald M. Paranal, Christopher B. Burge, James E. Bradner, Tong Ihn Lee, and Richard A. Young. 2012. "Transcriptional Amplification in Tumor Cells with Elevated c-Myc." *Cell* 151 (1):56–67. https://doi.org/10.1016/j.cell.2012.08.026.

Lonsdale, John et al. 2013. "The Genotype-Tissue Expression (GTEx) Project." *Nature Genetics* 45 (6):580–85. https://doi.org/10.1038/ng.2653.

Lovén, Jakob, David A. Orlando, Alla A. Sigova, Charles Y. Lin, Peter B. Rahl, Christopher B. Burge, David L. Levens, Tong Ihn Lee, and Richard A. Young. 2012. "Revisiting Global Gene Expression Analysis." *Cell* 151 (3):476–82. https://doi.org/10.1016/j.cell.2012.10.012.

Lovén, Jakob, Heather A. Hoke, Charles Y. Lin, Ashley Lau, David A. Orlando, Christopher R. Vakoc, James E. Bradner, Tong Ihn Lee, and Richard A. Young. 2013. "Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers." *Cell* 153 (2):320–34. https://doi.org/10.1016/j.cell.2013.03.036.

McAllister, James W. 1997. "Phenomena and Patterns in Data Sets." *Erkenntnis* 47 (2):217–28. https://doi.org/10.1023/A:1005387021520.

McAllister, James W. 2011. "What do Patterns in Empirical Data Tell Us About the Structure of the World?" *Synthese* 182 (1):73–87. https://doi.org/10.1007/s11229-009-9613-x.

Nie Zuqin, Gangqing Hu, Gang Wei, Kairong Cui, Arito Yamane, Wolfgang Resch, Ruoning Wang, Douglas R. Green, Lino Tessarollo, Rafael Casellas, Keji Zhao, and David Levens. 2012. "c-Myc is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells." *Cell* 151 (1):68–79. https://doi.org/10.1016/j.cell.2012.08.033/

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press

Rasmussen, Nicolas. 1993. "Facts, Artifacts, and Mesosomes: Practicing Epistemology with the Electron Microscope." *Studies in History and Philosophy of Science Part A* 24 (2):227–65. https://doi.org/10.1016/0039-3681(93)90047-N.

Schaum, Nicholas, et al. 2018. "Single-Cell Transcriptomics of 20 Mouse Organs Creates a *Tabula Muris*." *Nature* 562 (7727):367–72. https://doi.org/10.1038/s41586-018-0590-4.

Tulodziecki, Dana. 2007. "Breaking the Ties: Epistemic Significance, Bacilli, and Underdetermination." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 38 (3):627–41. https://doi.org/10.1016/j.shpsc.2007.06.002.

Tulodziecki, Dana. 2017. "Underdetermination." In *The Routledge Handbook of Scientific Realism*, edited by Juha Saatsi, 60–71. New York: Routledge.

Woodward, Jim. 1989. "Data and Phenomena." *Synthese* 79:393–472. https://doi.org/10.1007/BF00869282.

Wylie, Caitlin D. 2019. "Overcoming the Underdetermination of Specimens." *Biology & Philosophy* 34:24. https://doi.org/10.1007/s10539-019-9674-2.

Zhang, Hong, Fei Zhang, Yiming Yu, Li Feng, Jinbu Jia, Bo Liu, Bosheng Li, Hongwei Guo, and Jixian Zhai. 2020. "A Comprehensive Online Database for Exploring ~20,000 Public *Arabidopsis* RNA-Seq Libraries." *Molecular Plant* 13 (9):1231–33. https://doi.org/10.1016/j.molp.2020.08.001.

Cite this article: Tulodziecki, Dana. 2025. "Data Can Be Underdetermined, Too." *Philosophy of Science* 92 (5):1500–1510. https://doi.org/10.1017/psa.2025.10147

https://doi.org/10.1017/psa.2025.10147 Published online by Cambridge University Press