



Historical Case Studies: The “Model Organisms” of Philosophy of Science

Samuel Schindler¹ · Raphael Scholl²

Received: 26 March 2019 / Accepted: 29 January 2020 / Published online: 14 February 2020
© Springer Nature B.V. 2020

Abstract

Philosophers use historical case studies to support wide-ranging claims about science. This practice is often criticized as problematic. In this paper we suggest that the function of case studies can be understood and justified by analogy to a well-established practice in biology: the investigation of model organisms. We argue that inferences based on case studies are no more (or less) problematic than inferences from model organisms to larger classes of organisms in biology. We demonstrate our view in detail by reference to a case study with a long history: Semmelweis’s discovery of the cause of childbed fever.

1 Introduction

Case studies are ubiquitous in the philosophy of science: they can be found in classical debates concerning theory appraisal, scientific realism, explanation, and many more. Crucially, case studies are usually taken to be representative of a broader class of cases. For example, Lakatos took his discussion of development of the Bohr model of the atom to be representative of research programmes *in general*. When Kuhn argued that the meaning of the term ‘mass’ is different in Newtonian and relativistic mechanics, Kuhn believed to have lent support to the thesis that paradigms *in general* are incommensurable. More recently, scientific

Samuel Schindler and Raphael Scholl have contributed to this paper equally and appear alphabetically.

✉ Samuel Schindler
samuel.schindler@css.au.dk

Raphael Scholl
ras223@cam.ac.uk

¹ Center for Science Studies, Department of Mathematics, Aarhus University, Ny Munkegade 118, 8000 Aarhus, Denmark

² Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge CB2 3RH, UK

realists have argued on the basis of historical cases like the Fresnel wave theory of light that realism as a *general* position about science is sustainable despite radical theory change.¹ Similarly, the debate about whether the Hodgkin–Huxley model is explanatory is believed to have implications not only for this particular model, but for a large share of explanations in biology *in general*.²

The case study approach has long been criticized on a number of grounds. In his famous “marriage of convenience” paper on the difficult relationship of the history and the philosophy of science, Giere (1973) criticized the approach as being “without a conceptually coherent programme” because it did not address the question of how historical facts can support philosophical norms. Nickles (1995) summarized a widespread sentiment when he wrote that “historical case studies can be too much like the Bible in the respect that if one looks long and hard enough, one can find an isolated instance that confirms or disconfirms almost any claim” (141). That is, case studies may be theory-laden by the philosophical claims they are supposed to support and may be cherry-picked to support those claims (Schickore 2011; Kinzel 2015; Nickles 1995; Hull 1992; Pitt 2001). Furthermore Pitt (2001) worried that “it is unreasonable to generalize from one case or even two or three” and believed that case studies run the risk of being “manipulated to fit the point” (373). If these criticisms are correct, then case studies cannot support even moderately general philosophical conclusions.

Several contributions have sought to address these concerns. Giere and Laudan, for example, have argued that the norm–fact divide can be overcome by a naturalism that views norms as instrumental norms (Giere 1985, 1989; Laudan 1987, 1990, 1986). It has also been suggested that although historical facts can motivate the construction of norms, historical facts are not what justifies philosophical norms (Schindler 2013). Kinzel (2015) argued that for several important evidential functions of case studies, the problem of theory-ladenness in the history of science is of no more concern than it is in the sciences. Scholl and Rüz (2016) proposed that cherry-picking can be addressed by explicating one’s selection criteria, and Scholl (2018) argued that worries about cherry-picking rely to some extent on a mistaken conception of the method of integrated HPS.

The problem of extrapolating from case studies has not been tackled as directly as the other issues. One approach understands the method of history and philosophy of science as analogous to hypothetico-deductive theory-testing, where general philosophical claims are tested against particular historical facts (Donovan et al. 1988; Laudan et al. 1986). Many recent commentators, however, have resisted thinking about the history-philosophy relationship in terms of theory testing. Chang (2011), Schickore (2011) and Currie (2015), for example, have sought to conceive the relationship between history and philosophy as, respectively, a relation between the concrete and the abstract, as a hermeneutic circle, or as concept illustration. Similarly, Lennox (2001) and Schickore (2011) have argued that instead of treating historical case studies as supporting philosophical theses, historical cases

¹ See Worrall (1989). For lists of historical cases that have been discussed in the realism debate see Laudan (1981) and Vickers (2013).

² See Weber (2008), Craver (2008) and Levy (2013).

can be used more fruitfully to study the historical origins and the development of philosophical issues.

Here we propose a new account of the role of case studies in the philosophy of science. The core idea is that historical case studies function analogously to a well-established practice in biology: the investigation of model organisms. We submit that the extrapolation from case studies to broader philosophical claims can be understood in the same way as the extrapolation from model organisms to broader biological claims. In biology, such extrapolation is usually taken to be warranted by phylogeny: the fact that populations and species are related by descent can ground our extrapolations in facts about similarity. We will outline an analogous phylogenetic justification for extrapolation from historical case studies.³ These inferences, as we will point out, are supported by two other features that historical case studies and model organisms share: many pragmatic factors determine the choice of a certain model organism or case study, and both model organisms and case studies are used repeatedly. Both of these features ease extrapolatory inferences and help us make progress in our understanding of the relevant science.

The account we will outline is to some extent intended descriptively, since we believe it captures many aspects of how case studies are already used with great success. However, it is also proposed in a normative spirit, since a keener awareness of the methodological issues that are involved in case study methodologies may improve future scholarship.

In Sect. 2 we review the philosophical literature on model organisms with an emphasis on their epistemic role. In Sect. 3 we spell out the analogies between case studies and model organisms concretely, by reference to a widely-used case study in the history and philosophy of science: Semmelweis’s investigation of the cause of childbed fever. In Sect. 4, we discuss the assumptions and implications of our approach in more detail, and we defend it against criticisms. Section 5 concludes our discussion.

2 Model Organisms and Case Studies

Biological model organisms are intriguing scientific objects: they promise inferences from a very limited set of instances to an indefinite one. As Ankeny and Leonelli (2011) write, “model organisms are always taken to represent a larger group of organisms beyond themselves” (318). Perhaps most famously, Thomas H. Morgan and his research group laid the foundation of modern genetics with their experiments on *Drosophila melanogaster* in the early twentieth century. But both before and after Morgan, many other organisms were established as models in particular fields and for particular research questions: sea urchins in early developmental biology, *Escherichia coli* in the study of bacterial growth and conjugation, squid for the study of nerve cells, mice for the study of the immune system, baker’s yeast in the study

³ Like Lennox (2001), we refer to our proposal as a “phylogenetic approach”, but we take the analogy to phylogenetic reasoning in biology in a different direction.

of eukaryotic cells, *Caenorhabditis elegans* for the study of the molecular basis for behavior and development, and *Arabidopsis thaliana* for the study of plants. Model organisms define the landscape of experimental biology.

However, it is striking just how limited the number of model organisms is in biology. As Weber (2004) put it felicitously, “molecular biology laboratories are extremely impoverished in biodiversity” because “most laboratories work on only a single species, and a large number of laboratories work on the same species” (155). Weber suggests three related questions about this seemingly peculiar practice of using model organisms: (i) why do biologists choose particular species as their model organisms?, (ii) why do biologists keep using the same model organisms instead of diversifying their induction base?, and (iii) how is it possible to extrapolate from model organisms to other organisms such as humans?

With regard to question of why biologists choose particular model organisms, Weber suggests that pragmatic reasons play an important role in addition to epistemic ones. For example, the organism must be easy to breed in the laboratory, its generation time must be short, and its features must be suitable for specific research questions (e.g., the size of the squid giant axon, or the size of chromosomes in *Drosophila*'s larval salivary glands, 176ff.). With regard to the question of why biologists return to a limited set of models, Weber argues that standardization has a positive cumulative effect. Once experimental techniques and procedures have been developed, it is reasonable not to shift to different organisms where the known experimental techniques might not work as well and where new techniques might have to be developed (175f.). A closely related advantage of returning to known model organisms is that this makes it easier to reproduce results and to build on them step by step.

With regard to the question of extrapolation, Weber argues that inferences from model organisms to other organisms are grounded in phylogeny, that is, in their evolutionary history (180f.). An inference from a model organism (such as fruit flies) to a target organism (such as humans) is thus justified because both the model organism and the target organism share features inherited from a common ancestor. Levy and Currie (2014, 333) develop the phylogenetic grounding of model organisms in more detail. In particular, they distinguish theoretical modelling from “empirical extrapolations” involving model organisms. They argue that in theoretical modelling, one must always check whether the target is actually similar in the relevant aspects to the model in order for the model-inferences to be justified. By contrast, in empirical extrapolations involving model organisms, “the relatedness of the lineages licenses inferring from one to another, *without the need to explicitly compare the underlying traits*” (330, our emphasis). Inferences on the basis of model organisms, according to them, are thus justified (implicitly) by phylogeny.

We believe that the use of historical case studies in the philosophy of science can be understood along the lines laid out by Weber for model organisms in biology. That is, (i) historical case studies are selected in part for pragmatic reasons such as simplicity, comprehensibility, and ease of investigation, (ii) they are used repeatedly in part because it is efficient for philosophers to re-use the relevant resources and to build upon previous analyses, and (iii) case studies allow extrapolation to broader

classes of cases because each case is embedded in scientific research traditions which relate it to other cases.

At this point, we hasten to offer two caveats. First, it is well documented that biological practices centered upon model organisms have drawbacks and limitations. For instance, Bolker (2014, 2017) has argued compellingly that model organisms may be an efficient tool of investigation in some cases but not in others, and interest in “non-model” model organisms has increased significantly since Weber’s writing more than a decade ago (see, e.g., Russell et al. 2017). Whether standard model organisms are suitable will depend on the question to be asked. *Drosophila* and *C. elegans* are excellent models for the genetic control of development in part because their rapid development is highly canalized, that is, because it is less responsive to environmental variation than the development of other organisms. The flip side of this is that neither *Drosophila* nor *C. elegans* are first-line choices for studying phenotypic plasticity. Similar considerations apply in the philosophy of science: Whether a case study is suitable for studying a particular question will depend on the question.

Second, phylogeny may provide a reason for expecting some results obtained in a model to extrapolate to others, but this expectation will be disappointed in some instances. In both biology and the history and philosophy of science, results obtained in a model must be confirmed in the target if we demand certainty. We are offering phylogenetic relationships as respectable grounds for extrapolation, not as a guarantee that extrapolation will be successful. Indeed, it can be argued that one of the hallmarks of generalization in biology is that both similarities *and* differences will be expected between model and target organisms (Bechtel 2009).

In the next section, we will flesh out our account by considering in detail one of the classic case studies in the philosophy of science: Ignaz Semmelweis’s discovery of the cause of puerperal fever between 1844 and 1848. We will structure our discussion along the three questions raised by Weber about model organisms.

3 Choosing, Stabilizing and Learning from a Case Study: Semmelweis on Puerperal Fever

Semmelweis’s discovery of the cause of puerperal fever began its life as a philosophical case study in Hempel’s *The Philosophy of Natural Science* (1966). Hempel explained that Semmelweis, a physician working in Vienna around the middle of the nineteenth century, was motivated by a puzzle. The mortality rate of so-called childbed fever differed markedly between two divisions of the same maternity clinic: in the first division, the mortality rate was near 10%, but in the second division it was comparatively low at 3%. On Hempel’s account, Semmelweis demonstrated the cause of the difference using the hypothetico-deductive method. He framed a number of hypotheses to explain the difference: He suspected differences in weather conditions, hospital crowding, birthing positions, and examination techniques, among others. But he found that each hypothesis yielded false predictions. Eventually Semmelweis hit upon a more successful hypothesis.

The first division was run by physicians, who conducted autopsies before examining pregnant patients, while the second division was run by midwives, who performed no autopsies. Semmelweis surmised that the physicians transferred some kind of infectious substance from autopsies to patients. This hypothesis yielded correct predictions since the institution of thorough hand-washing measures reduced the mortality in the first division to levels below those of the second. On the hypothetico-deductive reconstruction, we can understand this as cycles of conjecture and refutation followed by an eventual confirmation.

In the succeeding decades, the Semmelweis case was revisited by numerous historians and philosophers of science. While the case is often treated as a mere toy model, several philosophers have used it as a serious object of study that provides insight into scientific reasoning. The most extended treatment was by Peter Lipton (2004), who used Semmelweis as the main case study in his book *Inference to the Best Explanation*. Lipton aimed to show in detail how Hempel's hypothetico-deductive account fails, while inference to the best explanation succeeds, at capturing and justifying Semmelweis's actual scientific reasoning. Another important contribution was by Gillies (2005), who studied Kuhnian factors in the case in order to explain why Semmelweis's findings were initially rejected. Bird (2010) argued that Semmelweis's reasoning should be understood as an instance of inference to the *only* explanation. Later writers continued to find new aspects in Semmelweis's reasoning. Scholl (2013, 2015) found Semmelweis's inferences to correspond closely to J. S. Mill's (1843) methods of experimental inquiry and argued that an explanationist framework was not necessary to recover Semmelweis's inferences. Taking a different tack, Tulodziecki (2013) argued that Semmelweis's reasoning was often careless and should not be held up as a paradigm of scientific inference at all.

Our contention is that practices centered on historical case studies can be understood by analogy to practices centered around model organisms. To see that this analogy is helpful, we will ask Weber's three questions about model organisms about the Semmelweis case: Why was the case considered suitable in the first place? Why have philosophers and historians of science returned to it repeatedly instead of considering other cases? And why do they think that concepts that are useful for understanding the Semmelweis case speak to the question of scientific discovery and confirmation more broadly?

3.1 Why are Particular Episodes Chosen as Case Studies?

Case studies, like biological model organisms, are chosen in part for epistemic and in part for pragmatic reasons. We will have much more to say about epistemic reasons in later sections. Here we will focus on pragmatic reasons.

Looking at biology, *Drosophila* became a preferred model organism because it is easy to breed, has a short life cycle, and is rich in phenotypically traceable mutants (Weber 2004, 177). Likewise, philosophers select certain historical case studies because they are straightforward to present and to understand (they offer cognitive ease), and because they are rich in philosophically informative detail. Hempel, for

example, chose the Semmelweis case in part because it is “a simple illustration of some important aspects of scientific inquiry” (1966, 3).

Cognitive ease comes with risks. Cases that are easy to present and understand are not necessarily typical of many aspects of science—just as the plant *Arabidopsis thaliana* is relatively quick and easy to breed but perhaps not representative of many aspects of the long-lived *Sequoia sempervirens*. Importantly, however, atypicality need not be an obstacle to inductive reasoning. Some atypical traits are unrelated to the traits under investigation. A high breeding rate need not distort Mendelian ratios or impede the production of chromosomal maps. Even when atypical traits themselves are investigated, this need not be a problem for extrapolation. For example, the discovery of the mechanisms of action potential propagation was enabled by the giant squid axon, which, as its name suggests, is atypically large. But circumstantial evidence indicated early on that the same mechanisms are shared by more typically sized nerve cells (Levy and Currie 2014, 334f.).

The situation is similar for the Semmelweis case. Some of its atypical features are irrelevant to the philosophical research questions it has served to illuminate. An example of this is the case’s eventual appropriation by Hungarian nationalism: This is atypical in the sense that most medical discoveries we are interested in were not appropriated in this way, but it is also irrelevant to the practices of discovery and confirmation in medical science half a century before that appropriation. Other atypical features of the case, by contrast, are relevant to our philosophical concerns. A didactically appealing feature of the case is that it begins with a striking difference in mortality between two distinct hospital divisions, a contrast which immediately suggests the search for differences between the wards to account for it. Lipton, for instance, has argued that this illuminates the way in which scientific hypotheses are generated more generally (2004, 73). But is extrapolation not impeded by the very fact that the two divisions offered such a stark and unusual contrast? We do not think so. The Semmelweis case merely offers an accentuated version of the kinds of unexplained contrasts that often drive research. More typically, relevant initial contrasts simply occur in mixed populations rather than neatly pre-sorted into hospital divisions. In this sense, the contrast that motivated Semmelweis’s research is merely an exaggerated version of a common trait, like the squid giant axon. Thus, in spite of its atypical features, the Semmelweis case may be highly representative of a broad range of cases of successful scientific hypothesis generation.

Another requirement for a good model organism is that it must present insightful variation (Weber 2004, 177). The use of *Drosophila* for genetic analysis was helped by the fact that its populations included mutants with phenotypic effects that were reasonably easy to discern. Similarly, case studies need to be suitably varied and complex to allow for interesting philosophical analysis. Semmelweis’s many hypotheses instruct us about both discovery *and* justification, experiment *and* observation, refutation *and* confirmation, and so on. Hempel’s discussion can thus touch on a broad range of issues: that narrow inductivism is false (because unobservable entities are by necessity postulated), that some hypotheses are refuted by observation and some by experiment (but that this is logically the same), and that even successful theories are fallible (because of the fallacy of affirming the consequent). Lipton similarly relied on the case’s richness when he argued that the

Semmelweis case gives us a handle on the nature of hypothesis generation (since we search for explanatory differences between contrasting groups) and that it elucidates why some explanatory failures speak against a theory while others are simply irrelevant to it (it depends on the relative explanatory strengths of competing hypotheses). The Semmelweis case is thus suitable in part because of its richness. It allows a wide range of philosophically interesting questions to be raised and examined, just as *Drosophila* offered variation in many genes and many traits.

In sum, we see the choice of a historical case study as analogous to the choice of model organisms: the goals are fundamentally epistemic, but the selection criteria are partly pragmatic. It matters whether the case offers cognitive ease, whether its atypical features are relevant to our research question, and whether it is sufficiently rich to provide answers to a whole range of interesting questions.

3.2 Why are Case Studies Used Repeatedly?

We saw in Sect. 2 that there are at least two good reasons for reusing the same model organism instead of expanding our set of cases: standardization and reproducibility. Once laboratory techniques have been adapted to a particular model organism and have become productive, it is costly to switch to a different experimental system. What is more, switching may make it harder to compare, contrast and integrate results from different studies. If our goal is to study a particular biological system in depth, it makes no sense to switch exemplars.

The reuse of historical case studies relies on similar considerations. Once editions of source documents and other secondary works have been written, it is efficient to keep studying the now well-documented episodes. Starting anew requires disproportionate effort. Because of this, case studies can become entrenched, much as well-understood model organisms become entrenched. Standardization is both a powerful tool and a cause of inertia, in both disciplines.

Even if resources were unlimited, however, an argument can be made for reusing case studies. Returning to the same case repeatedly allows our analyses to build on each other in a cumulative manner. We will see below that there is much to be learned by comparing different reconstructions of the *same* episode (Sect. 4.1). Although this is not precisely analogous to natural scientists' concern for reproducibility, it serves some of the same functions. After all, reproducibility in science is not only of interest because it allows us to double-check our results. It moreover permits us to investigate an established phenomenon further: the precise conditions under which it occurs, the intermediate steps by which it is realized, and so on. In biology as in HPS, established phenomena provide scaffolding for further work.

The Semmelweis case illustrates the advantages of standardizing case studies: the relevant sources and background materials are easily available and often edited, so that research on conceptual questions can proceed from a rich foundation. By the time of Lipton's (2004) use of the Semmelweis case as an extended study of inference to the best explanation, Semmelweis's main work, the *Etiology, Concept and Prophylaxis of Childbed Fever*, had already appeared in a new and accessible

English translation by K. Codell Carter (Simmelweis 1983). The translation had been written expressly in order to facilitate philosophical and historical study of the case, particularly in the context of introductory courses in the philosophy of science (see Carter’s introduction to the translation). Most writers from then on used the new translation: both the “Kuhnian” take on Semmelweis by Gillies (2005) and the “Holmesian” take by Bird (2010) rely on it. Carter also provided further historical material on Semmelweis, his work, and his predecessors, which proved valuable for the continuing study and reassessment of Semmelweis’s reasoning. Thus, a fair amount of research on the historical sources and the context of the Semmelweis case contributed to its standardization and further use as a case study.

However, the Semmelweis case also teaches the dangers of standardization. Carter’s translation of Semmelweis’s *Etiology* made editorial choices that reflected Hempelian preconceptions about Semmelweis’s goals and methods (Scholl 2013). For instance, many pages of numerical tables are left out of the translation because they appear repetitious from the hypothetico-deductive point of view. In truth, however, the numerical tables attest to Semmelweis’s use of methods akin to Mill’s methods of agreement and concomitant variation. Without these tables, the methodological core of the work is obscured. Similarly, Carter omitted an account of animal experiments, which do not feature in the hypothetico-deductive account. Yet some of Semmelweis’s contemporaries considered these animal experiments to be among the clinching evidence for Semmelweis’s case. Such omissions in secondary works can obviously reinforce existing biases.

To sum up, reusing case studies allows researchers to rely on existing resources such as editions and secondary literature. Moreover, the repeated use of the same case study enables convenient comparison between contrasting or complementary philosophical accounts.

3.3 How Do We Learn from Individual Case Studies?

We now come to what is perhaps the most interesting question concerning case studies: how can single cases enable extrapolation to a broader group of cases? In Sect. 2 we saw that inferences from model organisms to a broader group of organisms are justified phylogenetically. Mechanisms in *Drosophilidae* may sometimes reflect mechanisms in *Elephantidae* because some of the biological mechanisms of these two separate families are shared due to common descent.

What is the equivalent of phylogeny for case studies? It is historical influence: any episode which we isolate in the form of a case study has relations to research practices and traditions before and after. Researchers learn from and imitate each other. They pick up ideas from their colleagues and predecessors, modify them, and pass them on. Even innovative findings rest on such a foundation. Semmelweis’s discovery may have been a breakthrough for our understanding of infectious diseases, but its methodology is continuous with earlier work. Even in Semmelweis’s time, the creation of control groups and the exclusion of confounders were common concerns in clinical research. We find them, for instance, in the work of James Lind

in Britain or P. C. A. Louis in France (Matthews 1995; Tröhler 2000). Although many of the details of the methods for experimental causal inference were in flux, these researchers contributed to a shared methodological tradition. Semmelweis is interesting to us in part *because* he is a representative of this broader methodological tradition. Arguably, even today's randomized controlled trials are the offspring of the tradition in which Semmelweis worked.

The notion of historical connections between cases should be understood broadly. Beyond the lines of influence between major figures that historians of science used to be particularly interested in, broader methodological currents can be discerned, of which individual philosophers and scientists are representatives. They need not be links in a chain, for offshoots of a tradition (even terminal branches) can be informative.

The details of Semmelweis's causal inference methods provide a useful example. Scholl (2013) showed that Semmelweis's often abridged numerical tables reveal his reliance on inference methods that Mill characterized as his four canons of experimental inquiry. They include the methods of difference, agreement, concomitant variation, and residues. One way to think about this would be in terms of a "direct influence" hypothesis. Two influential methodologists articulated versions of the four methods of experimental inquiry in the first half of the nineteenth century: Herschel in his *Preliminary Discourse on Study of Natural Philosophy* of 1831, and Mill in his *System of Logic* of 1843. Mill was influenced by Herschel, according to his own statements, and to some extent merely systematized and named the methods outlined by the earlier author. Thus, the direct influence hypothesis is that Semmelweis must have read either Herschel or Mill for methodological inspiration. There is, however, no indication that Semmelweis read so broadly in the philosophy of science. At the same time, it is unlikely that Semmelweis's methodology is a case of independent invention, since his four methods match so specifically the ones outlined by Herschel and especially Mill. Some kind of historical influence is overwhelmingly likely. What seems most plausible to us is what we may call the "common roots" hypothesis. By the middle of the nineteenth century, the four methods were mainstays of scientific practice, and both Herschel and Mill successfully described and explicated these commonplace inference strategies. But the methods probably reached Semmelweis by different paths—for instance, by way of his Viennese teachers and their connection, among others, to Parisian exponents of the numerical method in the early nineteenth century.

The tradition of using versions of "Mill's methods" in experimental inquiry extended across disciplines, and Herschel, Mill, and Semmelweis were representative offshoots of it. What the phylogenetic approach brings out forcefully is that the interesting question is not just whether Semmelweis was directly influenced by Mill or Herschel, but how each of these historical actors was part of a growing body of scientific practices for inferring the causes of natural phenomena. By studying Semmelweis, we learn about these broader methodological currents.

In sum, one warrant for extrapolating from case studies is that individual cases are connected as parts of a branch of scientific thought and practice, just as model and target organisms are connected as parts of the tree of life. Because of their historical connections, studying one case can be expected to teach us something about others.

3.4 How the Three Functions Relate

In the previous three sections we explained that, just as in the practice of using model organisms in biology, (i) there are not only epistemic, but also pragmatic considerations going into the selection of historical case studies, (ii) case studies are used repeatedly in order to use resources effectively and to establish comparability, and (iii) historical relatedness justifies extrapolations from case studies. Function (i) and (ii) in fact facilitate extrapolation: some cases are especially suited for extrapolation because they exhibit certain features of scientific methodology in an easily accessible fashion (function i), and the concerted efforts of many philosophers working on a single case improves our chances of getting the induction base of our extrapolations right (function ii). Conversely, it would be much harder to establish reliable induction bases for inferences about scientific practice at large if the cases philosophers used were always extremely intricate and complicated, or if all philosophers used their own case studies, without consideration of the case studies used by other philosophers.⁴

4 Elaboration and Defense

We have argued that historical case studies in the philosophy of science work in close analogy to model organisms in biology. Cases are chosen for both epistemic and pragmatic reasons. The focus on a limited set of concrete cases allows us to study theoretically salient questions efficiently and incrementally. Researchers can reuse existing resources relating to the case and build cumulatively on each other's work to find out what, in the individual case, is involved in confirming hypotheses, explaining phenomena, reducing theories, or whatever aspect of science is being studied. So we can use case studies to hone our models of scientific practices.

In many cases, we will expect our results to extrapolate from the case under study to other cases. Just as in biology, extrapolation from individual exemplars typically depends on facts about historical relatedness. In biology, we can expect aspects of model organisms to match aspects of related species because they share common ancestors. In the case of history and philosophy of science, there is a comparable phylogenetic relationship. It is grounded in the transmission of concepts and practices between scientists. Individual scientists rarely invent out of thin air the standards by which theories are assessed, by which experiments are conducted, or by which explanations are constructed. The more usual situation is for such standards and practices to be transmitted by various routes from scientist to scientist. At least as a starting point of our investigation, there is thus good reason to expect that a detailed study of practices in one case will be transferable to some extent to other cases.

⁴ This is not to say, however, that complicated cases can never be used for extrapolatory inferences. On the contrary, sometimes it may be worth-while for the philosophical community to pick very complex cases and to direct all efforts on those. That would be advisable when methodological complexity is called for by the historical cases which the case study under consideration is supposed to elucidate.

In the present section, we will discuss both theses in more detail and support them by examples. First, what does the honing of our conceptual apparatus on the basis of a case study look like? Is this merely a descriptive exercise, or can repeated work on a case study have normative implications? Second, how precisely can we transfer the phylogenetic analogy to the history and philosophy of science, given that the mechanisms by which research traditions are formed, transmitted, and changed are presumably very different from the mechanisms by which biological traits are transmitted and changed?

4.1 Philosophical Progress by Historical Means

We have seen that Hempel's writings established Semmelweis's discovery of the cause of childbed fever as a paradigm case of discovery and confirmation. The case's pedagogical virtues helped to prompt a new English translation of Semmelweis's work by Carter in the early 1980s, and this set the stage for its deeper consideration by Lipton (1991, 2004) and a number of subsequent authors. Here we will focus on Lipton's treatment of the case. We will see that Lipton did not simply consider a broader range of historical facts than Hempel had, although this was certainly part of the project. Rather, he used the case to diagnose conceptual problems with Hempel's hypothetico-deductive account, and to suggest solutions to these problems in the framework of inference to the best explanation.

Lipton noted that one would expect the main difficulty of understanding scientific inference to lie in the justification of induction. Why are particular principles fit for purpose? In actual fact, however, the mere description of the principles by which we make inductive inferences has proved to be elusive. Lipton wrote: "[It] is not merely that we have yet to capture all the details, but that the most popular accounts of the gross structure of induction are wildly at variance with our actual practice" (2004, 12). He proceeded to show how, precisely, Hempel's hypothetico-deductive account was both too permissive and too restrictive to capture inductive inferences that Semmelweis actually made.

Consider, first, an example of the hypothetico-deductive account's excessive permissiveness. In the course of his investigation, Semmelweis rejected the hypothesis that childbed fever was caused by overcrowding of the hospital ward. Hempel reconstructed this in terms of a logical contradiction between the hypothesis and the observation that the two divisions were equally crowded. However, Semmelweis's rejection of the hypothesis would not have been licensed on hypothetico-deductive grounds, since the hypothesis that overcrowding causes childbed fever is perfectly compatible with equal crowding of the divisions. Overcrowding may cause childbed fever, for example, only in conjunction with other conditions that were only realized in the physicians' division. Overcrowding may also be only one of multiple causes of childbed fever, of which *others* were realized in the first division. Thus, argued Lipton, the hypothetico-deductive account cannot account for Semmelweis's rejection of the overcrowding hypothesis and is thus too permissive.

The second example, along similar lines, shows that the hypothetico-deductive account is also too restrictive, as it would not have licensed Semmelweis to accept

the hypotheses he actually accepted. Hempel took the cadaveric hypothesis to be confirmed by its observable consequences, in particular by the experiment showing that childbed fever decreased when hand-washing measures were instituted. If the cadaveric hypothesis is joined with appropriate auxiliary hypotheses, it may indeed entail the contrast between the experimental and the control group. However, there were many other relevant contrasts that the cadaveric hypothesis did not entail. While most women who delivered on the way to the hospital did not contract childbed fever, some did, even though they were far removed from autopsies. Similarly, even in the midwives' division, where no autopsies were performed, childbed fever occurred sometimes. The cadaveric hypothesis did not entail these contrasts. If it was justified to reject the overcrowding hypothesis simply because it did not entail the contrast between the physicians' and the midwives' divisions, then it would equally have been justified to reject the cadaveric hypothesis because it did not entail other salient contrasts.

Lipton made considerable headway on a descriptive track, by showing that the hypothetico-deductive account of confirmation cannot recover the inferences that Semmelweis actually made. Building on Hempel's previous writings, Lipton made a powerful case that the hypothetico-deductive account did not reflect actual scientific reasoning.

Critics have objected that Lipton's book suffers from the fact that it discusses none of its examples of IBE in any depth, except for the Semmelweis case (Norton in preparation). We agree with this criticism. It is not our project to justify an empirically impoverished philosophy of science. In our view, a successful HPS project, if it is to follow the biological model, must continuously balance the in-depth study of individual cases with the broadening of its empirical scope to further cases (cf. Bolker 2014, 2017). Our approach suggests that studying particular cases in-depth is a powerful strategy for improving our conceptual models of salient aspects of science. But this is not the end point of inquiry. Conceptual models, once honed on the basis of cases, are meant to be applied to additional cases, to assess their scope, to test their merits, and to refine them further.

In treating historical case studies as the model organisms of the philosophy of science, we have so far emphasized how cases enable descriptive progress. Accurate descriptions are a challenging and worthwhile goal in their own right, but they also lay the groundwork for further inquiry, including the traditional normative projects of the philosophy of science. Lipton judged that we lacked an adequate descriptive account of inductive inferences in science, and so he developed an improved account by using the Semmelweis case as a "laboratory", as it were, of such inferences. However, he also sought to justify the resulting account. There are multiple accounts of how precisely descriptive facts and normative claims relate to each other, and our account is noncommittal in this respect.

One way to view the relationship between philosophical norms and historical facts, and in particular progress in our philosophical understanding of science, is provided by Lakatos (1978). Lakatos is known (and despised) for suggesting that the "misbehaved" history be banned into the footnotes of philosophical analyses of science. We of course do not subscribe to this dismissive attitude towards actual history. Nevertheless we do believe that another one of Lakatos's ideas is valuable

for practitioners of HPS, in particular when it comes to evaluation of philosophical norms. Lakatos suggested that different philosophical “methodologies” could be assessed on the basis of how many historical facts they could accommodate in a rational way. For example, Lakatos (1978) argued that Popper’s falsificationism would by default deem parts of scientific practice irrational in which counterevidence did not automatically lead to the refutation of the theories in question. In contrast, Lakatos’s own account of “research programmes” made rational sense of such instances, because successful research programmes ought not to be given up (even when there is counterevidence for some of its predictions) unless better alternative programmes are available. Thus, Lakatos thought that some historical facts could indeed help us determine the right philosophical accounts of science, or at least the ones that are better than others.⁵ Importantly, such a view need not commit itself to normativity being grounded in the historical facts (see Schindler 2018).

We emphasize that we are not necessarily committed to Lakatos’s account of progress in HPS. There are other accounts, such as Laudan’s normative naturalism, which view the relationship between philosophical norms and facts rather differently, and which nevertheless may provide solutions to the problem of progress in HPS.⁶

4.2 Transferring the Phylogenetic Analogy to the History of Science

The previous section discussed historical case studies as productive laboratories of philosophical inquiry. But we have also suggested that we can draw conclusions from the study of cases that extend to further cases. Because scientists work in research traditions, the practices observed in one part of the research tradition will often generalize to other parts of it. We compared this to biology, where some traits of organisms that share common ancestors are preserved between organisms, species, and higher groupings. In both the biological and the philosophical case, we have argued, propinquity of descent can serve to justify extrapolation.

Analogies between biological and cultural change are often seen as problematic because the mechanisms of cultural change may be quite dissimilar to the mechanisms of biological evolution (for discussion, see Hull 1988; Fracchia and Lewontin 1999; Gray et al. 2007; Mesoudi et al. 2004; Godfrey-Smith 2012; Lewens 2015). It is therefore important for us to clarify what commitments this analogy entails, and what implications it has. This will show that our phylogenetic account, far from endorsing a strong notion of cultural evolution, relies on rather modest presuppositions.

The most important commitment of our view is to the existence of scientific concepts, methods, and practices that are transmitted with some fidelity within clusters of scientists. We call these clusters “research traditions”. If scientific beliefs and norms were so changeable as to show barely any continuity, then

⁵ For an in-depth discussion of this sometimes overlooked aspect of Lakatos’s work see Schindler (2018).

⁶ Whether Laudan’s account in fact accommodates progress is debated. See Worrall (1988), Laudan (1989), and Worrall (1989).

the phylogenetic approach, which presumes meaningful historical relations between cases, would not be applicable. Similarly, extrapolation between cases on the basis of historical relatedness would be illegitimate if scientists regularly discovered anew what the most successful standards are for confirming a theory, or what makes for an adequate explanation of empirical phenomena. This is again analogous to the biological case, where phylogenetic inferences are only possible if traits are homologous (inherited from a common ancestor), but not if they are homoplastic (that is, if they are similar for reasons other than common descent, such as independent adaptation to similar environmental demands). The phylogenetic approach is applicable only to those beliefs and norms that are widely shared due to historical diffusion, that is, that are homologies rather than homoplasies.

Cases in which practices are historically transmitted are our main concern here. However, there may be circumstances in which extrapolation from one case to another is legitimate even though it is not grounded in history. If we found (against our expectation) that many prevalent scientific practices are continuously reinvented rather than historically transmitted, then historical case studies could still be useful as a kind of model organisms. Let us stipulate that many scientific practices really are homoplasies in the sense that they are near-optimal and easily rediscovered solutions to common problems. Perhaps Mill's four methods of experimental inquiry are in some sense obvious, and were independently discovered by Semmelweis rather than learned from other researchers. If so, we would still learn much from studying the individual case in depth, since other cases are by assumption similar, even if their similarity is not grounded in historical relatedness. To use a simple biological analogy, what we learn from studying the aerodynamics of the wings of birds might be useful for understanding the wings of bats, even though these are independent adaptations for flight. If such cases of independent invention of scientific practices were found to be common, then a strictly phylogenetic justification for extrapolation would decrease in importance.

It is an empirical question how many scientific practices are historically transmitted rather than independently re-invented. However, the widespread occurrence of historical transmission is hardly in dispute. Many widely-read authors in the history and philosophy of science have highlighted the importance of what can be described as "adaptive radiations" in the growth of knowledge. These range from Kuhn's (1996) paradigms (in the sense of influential exemplars that set the standard strategies for solving particular kinds of problems) to Hacking's (1992) "styles of reasoning" (the experimental, the statistical, and so on).⁷

Beyond the necessity for historical transmission, the phylogenetic account is not committed to any particular mechanisms by which scientific norms and practices are transmitted, or by which research traditions originate, grow, and decline. In his *Science as a Process*, Hull (1988) outlined a selectionist model of scientific change. This may be correct, but we do not have the data to establish such a conclusion. The phylogenetic account can get by on much leaner assumptions. We agree with recent

⁷ Notably, Kuhn made an explicit analogy between evolution by natural selection and the growth of scientific ideas in Chapter XIII of *Structure*.

authors on cultural evolution that multiple levels of cultural evolutionary processes can be distinguished (Godfrey-Smith 2012; Lewens 2015). Godfrey-Smith, for example, distinguished between cultural evolution at the micro-, meso-, and macro-levels. At the micro-level, we may ask whether beliefs and practices spread by a process of “Darwinian imitation”, with cultural variants forming discernible parent-offspring lineages, or whether there is a network of influences that do not show such lineages. At the meso-level, we may ask whether there is a process in which successful variants (regardless of their micro-level dynamics) proliferate and provide many independent “platforms” for further improvements—that is, whether there is a process of cumulative cultural adaptation. At the macro-level, tree-like topologies of cultural lineages may be discerned regardless of whether a process of Darwinian imitation occurs at the micro-level, or whether cumulative cultural adaptation occurs at the meso-level. A “population genetics” of culture is not needed for our phylogenetic justification of extrapolation to be plausible. It is the topology of historical relatedness from a bird’s eye perspective, so to say, that matters to our account. We are committed to the existence of historical lineages that legitimate extrapolation, but not to particular assumption about micro- or meso-level processes at the “zoomed in” view.

A standard criticism against the notion that cultural change can be compared to evolution notes that the topology of cultural lineages, unlike that of biological lineages, is not tree-like. The assumption is that biological lineages do not rejoin after splitting, while in culture cross-lineage transfers are abundant—ideas and practices from one lineage can in principle diffuse to any other, and so the resulting topology is reticular. However, this sort of criticism does not undermine our account, for at least two reasons. First, as biologists discovered long ago, the basic assumption that biological evolution is marked by strict branching is mistaken (Hull 1988; O’Malley 2014). Outside of vertebrate evolution, in particular in the evolution of bacteria and plants, hybridization and cross-lineage borrowing are frequent. Thus, reticular topologies are no knock-down argument against applying evolutionary thinking to culture. Second, our account of extrapolation is simply not affected by the problem of reticular topologies. There is an interesting line of work in cultural evolution that wants to use current homologies to determine historical patterns of branching (e.g., to determine the phylogeny of languages on the basis of words shared between them). For such approaches, reticular topologies are a real (although not an insurmountable) obstacle (Gray et al. 2007). But our approach is not of this sort. Our project is *not* to use homologies in current scientific norms and practices to infer a historical pattern of branching. We assume that the patterns of historical relatedness will be determined largely by routine historical methods. Our claim is only that the fact of a historical relationship, established independently, can be used as a basis for extrapolation from one case to others. It does not matter whether the relevant historical relationships are part of a branching or a reticular pattern.

We are making a strong claim about the importance of historical knowledge to at least some work in the philosophy of science. If a philosopher claims that a detailed case study speaks to a broader class of cases, then the validity of this claim depends in part on facts about how the case under study is historically related to these other cases. Our account thus brings out a key but perhaps underappreciated

role for history in philosophy of science. It would be wrong, however, to think that philosophers and historians can simply divide labor—that philosophers can import, as it were, ready-made phylogenetic knowledge from historians of science. Instead, we see the task of determining the relevant phylogenetic relationships as one that philosophers and historians of science share. In many instances, the selection of representative case studies will require integrated philosophical and historical judgment. The re-emerging hyphenated historians-and-philosophers of science (Laudan and Laudan 2016) bring precisely this kind of judgment to the table.

At what resolution should we trace lineages of conceptual and practical influence? We think that only a pluralistic stance that adapts itself to different research questions can succeed. However, finer resolutions will often be more useful. A biological example is once more instructive. The *overall* phylogenetic relationship between current human and cow populations may not matter greatly if we wish to compare a *particular* aspect of these organisms—for instance, when we study bovine mitochondria in order to learn about human mitochondria. The question is only how much bovine and human mitochondria differ. The same is true in history and philosophy of science: If we wish to learn about experimental approaches from the Semmelweis case, we require a close kinship between Semmelweis's experimental methodology and the methodology deployed in the other cases we are interested in. However, there is no need for the cases to stand in a close relationship or to belong to the same broader research tradition with regard to *other* aspects. Perhaps Semmelweis was in many ways quite idiosyncratic and atypical, standing outside the main medical and biological research traditions of the middle of the nineteenth century. Nevertheless, we can extrapolate from his experimental methodology to other cases so long as the methodologies are appropriately related. Thus, the phylogenetic relationship between different historical cases is at the core of our views on the justification of extrapolation. However, the relevant phylogenetic relationship must be understood as fine-grained rather than coarse-grained: our approach is grounded in the study not of broad research traditions, but of lineages of methodologies, concepts and practices.

It is an empirical question whether a case study represents a broad or a narrow range of further cases. Lipton took inductive reasoning to be a broadly shared and reasonably uniform aspect of science. He thus understood the Semmelweis case as representative of almost all inductive inferences. This is comparable to inferences about the genetic code in biology, which is a broadly shared trait with little variation between different branches of the tree of life. Studying the genetic code in any organisms would be informative of virtually all others. But other traits are more local and more variable: The bones of the human hand are homologous only to other descendants of early tetrapods, and they vary considerably from bats (long and delicate, to form a wing) to dolphins (short and thick, to form a fin). More recent writers view Semmelweis's methods more like tetrapod bone homologies than like the genetic code. They take them to be representative of other instances of experimental causal reasoning in nineteenth-century biology and medicine, but not necessarily of all instances of inductive inference throughout science. Some phylogenetic inferences reach far because they concern widely shared and relatively constant traits, while others are more local.

One of our reviewers has asked whether our biologically inspired approach is compatible with the notion that science is progressive, given that evolution, to

which we analogize the growth of science, is not. Our approach is not committed on this question, which relates to the fine-grained details of the mechanisms of change. In biology, natural selection adapts organisms to changing environments, and so any change in the environment may alter or even reverse previous trends. Therefore, no overall goal towards which evolution progresses can be identified. The mechanisms by which scientists judge and retain successful concepts and practices may be quite different. They may well lead to the accumulation of those concepts and practices that contribute to one or several more or less constant overall goals, such as the accurate description of natural regularities. If so, the process could reasonably be described as progressive. But our approach is also compatible with the opposing view that science does not progress, contrary to appearances. The careful study of cases, and the extrapolation of findings to related cases, might help to identify progressive and non-progressive parts of science. Our approach is a method for studying such questions empirically. It does not prejudice them.

5 Conclusion

The justification for extrapolating the lessons learned in historical case studies rests on a material assumption about scientists and scientific practices. The assumption is that few scientists invent concepts, norms, methodologies or practices out of thin air. Usually such tools are transmitted by teaching and imitation, and we can learn something about entire lineages by studying representative episodes. We have argued that this is analogous to the way in which biologists justify inferences from model organisms to larger groups by appeal to phylogenetic relationships.

The analogy between model organisms and case studies extends beyond the surface. Not only is the justification for inductive inferences from cases similar to the justification for inferences from model organisms, but the approaches also share strengths and weaknesses. It is at first glance surprising that biologists focus on a limited set of model organisms instead of expanding their empirical basis. However, this restriction allows researchers to standardize both organisms and techniques so that rapid progress can be made in the study of particular systems. A similar procedure is typical of the history and philosophy of science, where intricate accounts of explanation, confirmation, and many other topics, are usually honed by reference to a small set of cases that receive sustained discussion. There is a danger in phylogenetic inference, since model systems or case studies may be atypical in crucial respects but not recognized to be so. In such cases, extrapolations may eventually prove to be erroneous. However, errors are amenable to correction in the course of further research, and often we will be called upon to deepen our analyses of individual cases, and to expand the set of cases we are studying. The result is philosophical progress by historical means.

On reflection, the similarities between historical case studies and model organisms should be unsurprising. Both biology and the history and philosophy of science have similar research objects: complex historical entities marked by branching and

reticular descent with modification. Naturally their empirical methods will converge to some extent.

References

- Ankeny, R. A., & Leonelli, S. (2011). What's so special about model organisms? *Studies in History and Philosophy of Science Part A*, 42(2), 313–323.
- Bechtel, W. (2009). Generalization and discovery by assuming conserved mechanisms: Cross-species research on circadian oscillators. *Philosophy of Science*, 76(5), 762–773. <https://doi.org/10.1086/605790>.
- Bird, A. (2010). Eliminative abduction: Examples from medicine. *Studies in History and Philosophy of Science Part A*, 41(4), 345–352.
- Bolker, J. A. (2014). Model species in evo-devo: A philosophical perspective. *Evolution & Development*, 16(1), 49–56. <https://doi.org/10.1111/ede.12056>.
- Bolker, J. A. (2017). Animal models in translational research: Rosetta Stone or Stumbling block? *BioEssays*, 39(12), 1700089. <https://doi.org/10.1002/bies.201700089>.
- Chang, H. (2011). Beyond case-studies: History as philosophy. In S. Mauskopf & T. Schmaltz (Eds.), *Integrating history and philosophy of science* (pp. 109–124). Heidelberg: Springer.
- Craver, C. F. (2008). Physical law and mechanistic explanation in the Hodgkin and Huxley model of the action potential. *Philosophy of Science*, 75(5), 1022–1033.
- Currie, A. (2015). Philosophy of science and the curse of the case study. In C. Daly (Ed.), *The Palgrave handbook of philosophical methods* (pp. 553–572). London: Palgrave Macmillan.
- Donovan, A., Laudan, L., & Laudan, R. (1988). *Scrutinizing science: Empirical studies of scientific change* (Vol. 193). Baltimore: John Hopkins University Press.
- Fracchia, J., & Lewontin, R. C. (1999). Does culture evolve? *History and theory*, 38(4), 52–78.
- Giere, R. N. (1973). History and philosophy of science: Marriage of convenience or intimate relationship. *British Journal for the Philosophy of Science*, 24(3), 282–297.
- Giere, R. N. (1985). Philosophy of science naturalized. *Philosophy of Science*, 52, 331–356.
- Giere, R. N. (1989). Scientific rationality as instrumental rationality. *Studies In History and Philosophy of Science Part A*, 20(3), 377–384.
- Gillies, D. (2005). Hempelian and Kuhnian approaches in the philosophy of medicine: The Semmelweis case. *Studies in history and philosophy of science part C: Studies in history and philosophy of biological and biomedical sciences*, 36(1), 159–181.
- Godfrey-Smith, P. (2012). Darwinism and cultural change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599), 2160–2170.
- Gray, R. D., Greenhill, S. J., & Ross, R. M. (2007). The pleasures and perils of Darwinizing culture (with phylogenies). *Biological Theory*, 2(4), 360–375.
- Hacking, I. (1992). 'Style' for historians and philosophers. *Studies in History and Philosophy of Science Part A*, 23(1), 1–20.
- Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs: Prentice Hall Inc.
- Hull, D. L. (1988). A mechanism and its metaphysics: An evolutionary account of the social and conceptual development of science. *Biology and Philosophy*, 3(2), 123–155.
- Hull, D. (1992). Testing philosophical claims about science. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*.
- Kinzel, K. (2015). Narrative and evidence. How can case studies from the history of science support claims in the philosophy of science? *Studies in History and Philosophy of Science Part A*, 49, 48–57.
- Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press. Original edition, 1962.
- Lakatos, I. (1978). *The methodology of scientific research programmes: Volume 1: Philosophical papers*. Edited by J. Worrall and G. Currie. Cambridge: Cambridge University Press.
- Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science*, 48(1), 19–49.
- Laudan, L. (1986). Methodology's prospects. *Philosophy of Science Association (PSA)*, 2, 347–354.
- Laudan, L. (1987). Progress or rationality? The prospects for normative naturalism. *American Philosophical Quarterly*, 24(1), 19–31.

- Laudan, L. (1989). If it ain't broke, don't fix it. *The British Journal for the Philosophy of Science*, 40(3), 369–375.
- Laudan, L. (1990). Normative naturalism. *Philosophy of Science*, 57(1), 44–59.
- Laudan, L., Donovan, A., Laudan, R., Barker, P., Brown, H., Leplin, J., et al. (1986). Scientific change: Philosophical models and historical research. *Synthese*, 69(2), 141–223.
- Laudan, L., & Laudan, R. (2016). The re-emergence of hyphenated history-and-philosophy-of-science and the testing of theories of scientific change. *Studies in History and Philosophy of Science Part A*, 59, 74–77.
- Lennox, J. G. (2001). History and philosophy of science: A phylogenetic approach. *História, Ciências, Saúde-Manguinhos*, 8(3), 655–669.
- Levy, A. (2013). What was Hodgkin and Huxley's achievement? *The British Journal for the Philosophy of Science*, 65, 469–492.
- Levy, A., & Currie, A. (2014). Model organisms are not (theoretical) models. *The British Journal for the Philosophy of Science*, 66(2), 327–348.
- Lewens, T. (2015). *Cultural evolution: Conceptual challenges*. Oxford: OUP.
- Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). London: Routledge.
- Matthews, J. (1995). *Quantification and the quest for medical certainty*. Princeton: Princeton University Press.
- Mesoudi, A., Whiten, A., & Laland, K. N. (2004). Perspective: Is human cultural evolution Darwinian? Evidence reviewed from the perspective of The Origin of Species. *Evolution*, 58(1), 1–11.
- Nickles, T. (1995). Philosophy of science and history of science. *Osiris*, 10, 139–163.
- Norton, J. D. (in preparation). *The material theory of induction*.
- O'Malley, M. (2014). *Philosophy of microbiology*. Cambridge: Cambridge University Press.
- Pitt, J. C. (2001). The dilemma of case studies: Toward a Heraclitian philosophy of science. *Perspectives on Science*, 9(4), 373–382.
- Russell, J. J., Theriot, J. A., Sood, P., Marshall, W. F., Landweber, L. F., Fritz-Laylin, L., et al. (2017). Non-model model organisms. *BMC Biology*, 15(1), 55. <https://doi.org/10.1186/s12915-017-0391-5>.
- Schickore, J. (2011). More thoughts on HPS: Another 20 years later. *Perspectives on Science*, 19(4), 453–481.
- Schindler, S. (2013). The Kuhnian mode of HPS. *Synthese*, 190(18), 4137–4154.
- Schindler, S. (2018). *Theoretical virtues in science: Uncovering reality through theory*. Cambridge: Cambridge University Press.
- Scholl, R. (2013). Causal inference, mechanisms, and the Semmelweis case. *Studies in History and Philosophy of Science Part A*, 44(1), 66–76.
- Scholl, R. (2015). Inference to the best explanation in the catch-22: How much autonomy for Mill's method of difference? *European Journal for Philosophy of Science*, 5(1), 89–110.
- Scholl, R. (2018). Scenes from a Marriage: On the confrontation model of history and philosophy of science. *Journal of the Philosophy of History*, 12(2), 212–238.
- Scholl, R., & Räs, T. (2016). Towards a methodology for integrated history and philosophy of science. In T. Sauer & R. Scholl (Eds.), *The philosophy of historical case studies*. Boston Studies in the Philosophy and History of Science.
- Semmelweis, I. (1983). The etiology, concept, and prophylaxis of childbed fever, translated by K. Codell Carter (Madison, WI: University of Wisconsin Press, 1983).
- Tröhler, U. (2000). *To improve the evidence of medicine. The 18th century British origins of a critical approach*. Edinburgh: The Royal College of Physicians.
- Tulodziecki, D. (2013). Shattering the myth of Semmelweis. *Philosophy of Science*, 80(5), 1065–1075.
- Vickers, P. (2013). A confrontation of convergent realism. *Philosophy of Science*, 80(2), 189–211.
- Weber, M. (2004). *Philosophy of experimental biology*. Cambridge: Cambridge University Press.
- Weber, M. (2008). Causes without mechanisms: Experimental regularities, physical laws, and neuroscientific explanation. *Philosophy of Science*, 75(5), 995–1007.
- Worrall, J. (1988). The value of a fixed methodology. *The British Journal for the Philosophy of Science*, 39(2), 263–275.
- Worrall, J. (1989). Fix it and be damned: A reply to Laudan. *The British Journal for the Philosophy of Science*, 40(3), 376–388.