Two Modes of Reasoning with Case Studies[1]

Wolfgang Pietsch[2], Munich Center for Technology in Society, Technische Universität München, Arcisstr. 21, 80333 München, Germany

I distinguish a predictive and a conceptual mode of reasoning with case studies. These broadly correspond with two different kinds of analogical inference, one relying on common and differing properties, the other on structural similarity. The problem of generalizing from case studies is discussed for both. Regarding the predictive mode, eliminative induction provides a natural framework. In the conceptual mode, general rules are largely lacking not least due to a number of epistemological challenges like Raphael Scholl's underdetermination problem for HPS. In agreement with ideas of Richard Burian and Peter Galison, I argue that conceptual reasoning on the basis of case studies should not aim at grand universal schemes but rather at mesoscopic or middle-range theory. In the essay, I will repeatedly draw on insights from the social sciences, in which a much more extensive reflection on case study methodology exists compared with HPS.

1. Introduction

By examining historical episodes in view of specific conceptual questions, case studies provide the essential link between the history and the philosophy of science. They are thus rightly regarded as the central building block of an epistemology for an integrated HPS, in short &HPS. From this perspective, it is rather surprising that there exists little systematic literature on this topic. Presumably, the reasons lie in certain methodological prejudices of both fields. On the one hand, philosophy of science often relies on toy examples or caricatures rather than engaging with the complex and intertwined details of actual historical episodes. In history, on the other hand, the neglect of a systematic reflection seems to stem from a wide-spread hostility towards abstract theorizing. But obviously, as long as historical narratives are framed in a language, their recounting will require a certain amount of conceptualizing.

In Section 2, a brief overview of some main characteristics of case studies will be given. Also, case studies will be contrasted with other research methodologies, in particular experiments and statistical approaches.

In Section 3, two modes of reasoning with case studies will be distinguished, an inductive mode aiming at prediction and an abstractive mode for concept development. I will argue that these two modes correspond to two different types of analogical inference, which have rarely been kept apart: on the one hand, inferences on the basis of corresponding and differing properties, on the other hand, inferences on the basis of structural similarity.

It turns out that there are related epistemic worries about the reliability of analogical inference and about reasoning with case studies. A considerable number of researchers believe that both

[1] Forthcoming in a collected volume on „The Philosophy of Historical Case Studies", edited by Tilman Sauer und Raphael Scholl (Boston Studies in Philosophy of Science)

[2] pietsch@cvl-a.tum.de

have only a heuristic value for suggesting plausible hypotheses. I will argue against this view in Section 4 and will show that the role of analogical and case-based reasoning is much richer. In particular, the problem how to meaningfully generalize from case studies will be addressed, both for the predictive and the conceptual mode. In line with most of the social science literature, I will argue that predictive inferences mainly rely on comparative reasoning based on the logic of eliminative induction in the tradition of Mill's methods. The situation is more complicated for the conceptual mode which is troubled by the ambiguities of concept formation. The crucial question, here, concerns the level of generality for the conceptual framework.

In a brief outlook, Section 5 provides further evidence for the need of a thorough methodological assessment of case-based reasoning across the sciences. With the recent emergence of data-intensive science, the boundary between statistical approaches and case studies becomes increasingly blurred – requiring a combined methodology for both. Currently, this has an impact mainly on the predictive mode and less on the conceptual mode.

Section 6 will summarize the findings with respect to the role of case studies in integrated history and philosophy of science. Especially when examining questions of scientific methodology, well-developed case studies are an indispensable part of the inquiry.

2. The nature of case studies

2a. General remarks

Notwithstanding their methodological importance, the nature and function of case studies is not a common topic in philosophy of science. In other fields, however, in particular the social sciences, there exists a sizeable and sophisticated literature on this issue, on which I will repeatedly draw in the following. While there are certainly substantial differences in the use of case studies compared with &HPS, there are also sufficient common grounds to merit a detailed look.

When defining the notion of case studies, the following cluster of criteria is usually mentioned:[3] (i) Case studies examine an episode in considerable detail and take into account the context in which it happened. Often, the examined phenomenon and its context are not clearly separable from each other. (ii) Relatedly, case studies take into account a large number of variables, while focusing on a single or at most a small number of instances. (iii) Often, the phenomenon of interest is examined from a variety of perspectives and with a variety of methods resulting in a heterogeneous data structure[4]. For example, case studies may involve empirical investigations, archival work, interviews, and surveys with open-ended questions. Triangulation is often used to verify that the different perspectives are actually coherent with each other. (iv) Importantly, case studies are always case studies *for* something – answering

---

[3] E.g. Yin 2009, p. 18; Gerring 2007, p. 17; George & Bennett 2005, p. 17-19.
[4] The term is understood here in a very broad sense, e.g. a historical narrative could also constitute a data structure.

the question 'what is this a case of?', i.e. they relate the examined episode to a certain theoretical concept or a type of phenomenon.

Some people apparently believe that case studies concern only a single, immutable instance, a snapshot of an event. But as for example John Gerring has pointed out, case studies always involve various kinds of within-case variation, for example temporal and spatial (2007, 27). In general, the more detailed a case study is, the more variation will occur. Also, case studies never stand entirely on their own, but are always contrasted at least implicitly with similar or dissimilar cases in the background knowledge. If nothing else, already the use of both ordinary and scientific language establishes the link with comparable instances.

Case studies are generally categorized as a type of qualitative research. However, especially when dealing with complex episodes, they may very well incorporate various quantitative elements, e.g. gradual variation of certain parameters over time or between different entities. Thus, the qualitative-quantitative distinction is not really suitable to characterize this type of research.

Relatedly, case studies are often contrasted with statistical methods. Mostly, a certain hierarchy is thereby implied that reliable inferences can only result from statistical reasoning and that the conduct of case studies is merely of heuristic value. However, this viewpoint is strongly and in my view rightly contested by most proponents of case study research (e.g. Yin 2009, p. 6). As already mentioned, case studies often involve statistical arguments. Even more importantly, a systematic analysis of within-case variation can lead to reliable analogical inferences as pointed out in Section 4b.

Finally, case studies are sometimes compared with experiments. Here, the latter in fact yield more reliable inferences because one can interact with and manipulate the phenomena while case studies mostly rely on observational data. Another difference is that experiments are usually carried out in a laboratory setting providing for a controlled and stable environment, while, as mentioned, the distinction between context and phenomenon is often helplessly blurred in case studies. However, the similarities between these types of research are more significant than the differences. Both case studies and experiments examine in detail a specific instantiation of the phenomenon of interest. Crucially, case studies share with certain classes of experiments their exploratory nature and the comparative logic of eliminative induction, which will be described in more detail in Section 4b.

A number of classification schemes have been proposed for further distinguishing different kinds of case studies, in particular (i) classification with respect to epistemological function, e.g. conceptual refinement, hypothesis generation, or causal analysis (cp. George & Bennett 2005, p. 75), or (ii) with respect to the representativeness of an episode. [5] As an example of the latter, Scholl and Räz offer a distinction between paradigm cases, hard cases and important cases that are of intrinsic interest.[6]

[5] A further distinction concerns the type of variation that occurs in a case study (e.g. Gerring 2007, p. 28).
[6] The terminology is employed in a talk, the slides of which are available online http://www.raphaelscholl.org/storage/talks/scholl-hps-22-11-13.pdf, accessed 25.11.2014. In the social-science literature, similar classification schemes can be found, e.g. Gerring suggests nine different types of cases:

For distinguishing different kinds of case studies, one can also have recourse to the comparative methodology underlying this type of research. 'Most similar' cases are as similar as possible in terms of circumstances, but the examined phenomenon fails to appear in one of the cases, allowing for the application of a weakened version of Mill's method of difference. 'Most different' cases are as different as possible, while the examined phenomenon still appears in both cases, allowing for the application of the method of agreement (e.g. Gerring 2007, pp. 89-90). In fact, already Francis Bacon in his well-known compilation of prerogative instances has listed a considerable number of useful types including 'solitary instances', which essentially correspond to 'most similar' and 'most different' cases, 'striking instances', which are paradigm or representative examples of a phenomenon, and the well-known 'instances of the fingerpost' or crucial instances (1620).

2b. Case studies in integrated history and philosophy of science

Very broadly, case studies provide the principal link between the factual level of the history of science and the conceptual level of the philosophy of science. One of the earliest attempts by James Conant, former president of Harvard University and mentor to Thomas Kuhn, to establish case-based reasoning in the history of science resulted in an influential book series *Harvard Case Histories in Experimental Science*. However, the idea that a few representative episodes can provide a complete picture, for example of the nature of experimental science has rightly been criticized in the aftermath – maybe most forcefully by Peter Galison in his *Image and Logic* on the grounds that there can be no unifying conceptual scheme because there exists no universal experimental method (1997, Ch. 1). Such and similar qualms were sometimes misinterpreted and have led to wide-spread skepticism about the usefulness of case studies among historians of science. However, Galison's point was the rejection of grand schemes and not an opposition towards any kind of conceptualization on the basis of case studies. Rather, he readily admits: "the very project of history must contend with the problem of how to make the particular stand for the general, and how to limit claims for such a stance" (Galison 1997, 60; see also Section 4c).

And in fact there is a long list of very successful examples in the history of &HPS, for instance Kuhn's study of the Copernican revolution (1992) which played a crucial role in his development of general concept like normal science or incommensurability, Hasok Chang's study of the development of a temperature scale (2004) that led to a general theoretical framework regarding the emergence of measurement techniques, or various analyses of Semmelweis' discovery of the importance of antiseptic procedures probing the adequacy of different kinds of scientific methodology (e.g. Hempel 1966, pp. 3-18; Scholl 2013).

3. The predictive and the conceptual mode of case-based reasoning

I will now distinguish two modes of reasoning with case studies, one aiming at the derivation of empirically verifiable predictions, the other at developing conceptual schemes that are useful to account for related phenomena. The first will henceforth be called the predictive

---

typical, diverse, extreme, deviant, influential, crucial, pathway, most-similar, and most-different (2007, pp. 89-90).

mode and the second the conceptual or abstractive mode. It should be obvious from the explanations that will follow that the boundary between both modes is not entirely sharp.

The distinction is situated within a broadly Duhemian or Cartwrightian view of scientific theories as having a phenomenological and a theoretical level (Duhem 1954, Ch. 2; Cartwright 1983, pp. 1-20). The former consists of causal laws that are often experimentally established. In accordance with their causal nature, these phenomenological laws are primarily used to generate successful predictions and interventions. By contrast, the theoretical level—more exactly, there are often several layers of increasing generality—consists of abstract or theoretical laws that are construed on top of the phenomenological level. These general laws are developed from the causal laws in a process of abstraction and are mainly geared at unification and explanation. While the causal laws can be true or false in a straightforward sense in that predictions based on them turn out either right or wrong, this is not so for the theoretical laws. In fact, a wealth of literature has established a solid case for underdetermination regarding the laws on higher levels of abstraction such as physical axioms. Thus, truth is not a suitable criterion to evaluate the theoretical level. Rather, pragmatic criteria like simplicity or fruitfulness are more adequate. I am aware that many aspects of this epistemological view could be questioned and I cannot defend it here except by referring to the authority of philosophers of science like those mentioned above.

Basically, the predictive mode of reasoning remains at the phenomenological level. It employs case studies to make predictions about other instances that are sufficiently similar, aiming at true or probable inferences either in terms of causal laws or of singular statements about future or unknown events. Mostly, such inferences do not change the level of description, i.e. the vocabulary remains at the same level of abstractness. Inferences in the predictive mode are mainly evaluated in terms of empirical adequacy.

The predictive mode is relatively rare in &HPS. To illustrate it, let us therefore look at an example from medicine. More than thirty years ago the first cases of what was later recognized as acquired immune deficiency syndrome (AIDS) began to appear in the Western world. Certain cancers and pneumonias that previously occurred only in patients with congenital immunodeficiencies or who had received immunosuppressive chemotherapy, were suddenly reported in patients with no such background (MMWR 1981). Confronted with a hitherto unknown condition, medical personnel were facing the severe challenge to develop both diagnosis and treatment. In such situations, where in-depth theoretical knowledge and understanding is missing, case studies both of single patients and of small clusters of patients are an adequate methodology to explore the phenomenon and to make predictions about similar cases. For example, the doctors measured various parameters of the immune system such as lymphocyte counts or T-cell counts. Also, the social background of the patients was examined with the result that almost all of them were gay and/or drug abusers and that the rate of sexual intercourse between them was much higher than would have been expected by pure chance, eventually suggesting the venereal character of the disease. Treatments were tried without much success and a large number of patients died.

Apparently, the detailed case studies of patients and small patient clusters were used as a reference to make predictions about further cases. They provided doctors with some information about what to expect and how to treat patients with similar characteristics. For example, if someone without a history of immunologic weakness presented him-/herself with infections that normally do not occur in patients with a healthy immune system, the doctors could fairly reliably guess in what range the various immunologic parameters will be or they might check the sexual partners of the patient to find further diseased individuals.

This fits well with the characteristics for the predictive mode as depicted above. In particular, the case studies of individual patients can be used for direct predictions about other instances that are sufficiently similar. Also, there is little conceptualization going on at this early stage of the discovery of AIDS, but rather the predictions are mostly framed using terms and concepts that were already employed for the collection of data.

By contrast, the abstractive mode aims at conceptual schemes. Abstractive reasoning usually changes the level of description, i.e. it employs a case study for the development of abstract laws on the theoretical level. To this purpose, it generalizes by selectively focusing on those aspects that are deemed essential within a specific research context. According to the epistemological perspective sketched at the beginning of this section, conceptualizations should not be judged in terms of truth but rather usefulness.

The majority of case studies in &HPS belong to this mode. A paradigm example was already mentioned in the previous section, namely Thomas Kuhn's development of notions like normal science or incommensurability based on detailed historical work in particular on the Copernican revolution (1992). On the basis of this case study, Kuhn proposed his view of scientific revolutions as reconceptualizations of a certain scientific area, often with broader implications for scientific methodology but also for the general philosophical world-view. The study already contains some of the core ideas that were later elaborated in Kuhn's more famous book on *The Structure of Scientific Revolutions* (1996). Kuhn suggests that there may be no final conceptualization for many fields of science and identifies non-cumulative elements in the process of scientific revolutions. In particular, while the range of phenomena that are accessible to explanation keeps growing, the nature of these explanations is subject to continuous changes during scientific revolutions (1992, 261-265).

The case study, which is of a very different kind compared with the example discussed for the predictive mode, nevertheless fulfills the criteria that were stated at the beginning of Section 2a. It illustrates well the characteristics of the conceptual mode. Obviously, concepts like a scientific revolution are situated on a much more theoretical level than the concrete historical events. Kuhn extracted the central features of such concepts by carefully abstracting from the complexities and contingencies of scientific practice. Propositions on the theoretical level, like "the paradigms before and after a scientific revolution are incommensurable", cannot be used to generate empirically verifiable predictions about what might be happening in a specific historical context. It therefore seems wrong to evaluate them in terms of truth. Instead, the crucial question is whether these concepts are useful for structuring the evidence in other historical episodes like for example the Darwinian or the chemical revolutions. While the notion of truth may apply when abstract concepts are used in concrete predictions on the phenomenological level, it generally cannot be employed to evaluate relations within the theoretical level. For example, Lamarckian concepts may fail to yield accurate predictions if applied to the question, why giraffes have acquired such long necks.[7] But this does not falsify Lamarckian ideas in general, it only shows a problem with a specific application of them.

Speaking of Kuhn, the use of case studies in &HPS bears much resemblance to that of exemplars in the natural sciences. Exemplars constitute representative cases that serve as role models for puzzle solving, they "are concrete problem solutions accepted by the group as, in a quite usual sense, paradigmatic." (Kuhn 1977, 298) Like case studies, they serve to develop a conceptual framework in order to analyze similar phenomena. The pendulum and the inclined plane are good examples, which in the hands of Galileo and others have contributed to establishing the conceptual foundations of classical mechanics (305-306). Another case in point is the harmonic oscillator whose principles can be understood by examining a simple physical system of a point-mass and a spring, but which is applicable in a large variety of research contexts far beyond physics.

Again, conceptual schemes based on exemplars are generally not used to predict but rather to adequately structure the perception of related phenomena. For example, the theory of harmonic oscillators does not itself allow for any predictions in a concrete context of application, e.g. in a resonant circuit. Rather, a number of concrete electrical laws must be presupposed, which should be formulated in a way that the theory of harmonic oscillators can be applied. Thus, theoretical knowledge is used to structure the more concrete knowledge regarding a certain context of application resulting in phenomenological laws and predictions which can then be verified or falsified.

Darwin also relied on paradigmatic phenomena in his discovery and construction of the theory of evolution, maybe most famously the variation of species on different islands of the Galapagos archipelago. Again, the fundamental concepts of evolutionary theory like mutation and selection are quite useful in coherently structuring phenomena in the world of living things, but they usually do not generate specific predictions. For example, they fail to determine whether the human species will become extinct in the next century or why the dinosaurs vanished from the planet.

In spite of these similarities, there are also important differences between the use of exemplars in the natural sciences and the conceptual mode of case-based reasoning in &HPS. (i) The first concerns the very nature of the concepts. While those developed in the natural sciences mainly regard the ontology of the natural world like forces or genes, case studies in &HPS mostly aim at methodological or occasionally sociological concepts like underdetermination or scientific revolutions. (ii) Another important difference concerns the historical nature of the studied phenomena. The exemplars examined in the natural sciences are generally repeatable and therefore largely independent of a historical context. By contrast, the episodes studied in &HPS occurred only once, mostly in a distant past. Trivially, one cannot intervene or experiment in historical case studies and the relevant evidence is usually not directly accessible, but has to be transmitted over time using media like books or artefacts. (iii) Finally, the complexity of the examined phenomena differs. While the exemplars in the natural sciences can often be separated from context and examined in a laboratory setting, this is not possible for case studies in &HPS.

[7] The example was suggested by Raphael Scholl.

Remarkably, the distinction between the predictive and the conceptual mode is mirrored in two different versions of analogical inference, which have not always been clearly kept apart. In one type, which corresponds to the predictive mode, an assessment of similarity in terms of common and differing properties determines if a further property of one case will be instantiated in another case as well (e.g. Mill 1886, Ch. XX, Keynes 1921, or Carnap 1980, Sec. 16&17). The other type is analogy in terms of structural similarity which has been employed since the ancients in accordance with the original meaning of the term 'analogy', referring to likeness in relations. Such structural analogies can be found over and over in the history of science and they are mostly employed for the conceptual development of novel phenomena. For instance, James Clerk Maxwell famously emphasized their importance for his work. A good example is his use of hydrodynamical and mechanical analogies for the formulation of the field-theoretic approach to electrodynamics. In the context of &HPS, Kuhn's conceptual framework of scientific revolutions is an example of structural analogies.

4. The problem of generalizing from case studies

4a. The problem stated

Attempts to generalize from case studies are often dismissed as naïve, calling into question the role of case studies as a sound piece of scientific methodology. In the following, I will briefly survey a number of arguments in this regard indicating where they go wrong and then tackle the problem of generalizing in a more specific way for the two modes of reasoning that were identified in the previous section.

(i) A wide-spread argument concerns the claim that one supposedly cannot generalize from a single immutable instance. But this misconstrues the notion of case studies, which as we had seen always include a considerable amount of variation, for example in time or across subjects if the case study involves more than a single subject or entity (Gerring 2007, pp. 27-33). A simple example of within-case variation that can generate predictions concerns interventions in a controlled environment that lead to sudden and substantial changes of a phenomenon. If an otherwise healthy person drinks a potion brewed from a previously unknown herb and dies quickly afterwards, the death is highly likely a result of consuming the herb and it is not advised that other people drink from the potion. The logic of this reasoning can be formally reconstructed in terms of eliminative induction, more exactly by the method of difference. Besides within-case variation, case studies sometimes involve cross-case variation by drawing comparisons with related cases. Finally, the argument overlooks that case studies are generally evaluated with respect to a substantial amount of background knowledge, also allowing for comparative reasoning.

(ii) Relatedly, the Humean view still prevails that one can only generalize from a large number of instances and that the level of verification is somehow proportional to the number of cases studied. For example, Joseph Pitt in an influential article on *The Dilemma of Case Studies* makes this point: "it is unreasonable to generalize from one case or even two or three." (2001, 373) This perspective is mistaken as the mentioned example of the poisonous potion shows. In a well-controlled environment, a causal connection can already be derived from two instances via Mill's method of difference, while a correlation even if established by a large number of instances may not mean anything. As argued by a number of methodologists in the tradition of Baconian eliminative induction[8], today often referred to as Mill's methods, it is not the number of instances, but the variation between instances that counts. Thus, if there is sufficient variation within a single case, we can generalize to other cases under similar circumstances. Especially in the social science literature on case studies, it is generally taken for granted that comparative reasoning based on Mill's methods is well-suited for the analysis of case studies though usually difficult to implement in full rigor (e.g. George & Bennett 2005, Ch. 8; Hammersley et al. 2000).

(iii) Many scientists regard case studies as merely anecdotal, serving at best a heuristic function for suggesting novel concepts and hypotheses. Certainly, some narratives that have been called case studies in the past are nothing but mere story-telling. However, the problem of differentiating a meaningful case study from an anecdote just amounts to developing a sound methodology for case-based reasoning. What is required is an analysis of the different types of variation that may occur and how they can lead to reliable inferences both in the predictive and the conceptual mode. A general framework is needed that provides an outline how to research, write, and analyze case studies. While social scientists have made important attempts at this task (e.g. Gary 2011, Gerring 2007, Yin 2009), in &HPS it still remains an underexplored question.

In the following, I will argue in further detail against the wide-spread view that case studies have a merely heuristic role in scientific method. I will again distinguish between the predictive and the conceptual mode.

4b. Generalizing in the predictive mode

As already indicated, inferences from case studies in the predictive mode mainly rely on an assessment of corresponding and differing properties between two instances. Note that such inferences are often made on a case-to-case basis and mostly do not result in sweeping generalizations. In the usual terminology, coined by Keynes (1921), the properties that two instances share are called the positive analogy, the properties that differ the negative analogy. Furthermore, it is useful to distinguish the known positive or negative from the unknown analogy. On the basis of the known positive and negative analogy, P and N, respectively, a probability is determined that a property q which is present in one instance can be found in another instance as well. Many methodologists seem to believe that such analogical reasoning cannot be based on a systematic general framework (e.g. Norton 2011). But while I would agree that a definitive account has yet to be developed, I am quite optimistic that it is possible and will provide a broad outline below.

In the 20th century, two influential programs have tried to tackle the task, one developed by John Maynard Keynes in his *Treatise on Probability* (1921), the other by Rudolf Carnap in

[8] The term 'eliminative induction' for Mill's methods has been used by several authors, in particular by Mill himself (1886, 256) and also by Mackie in his influential book *The Cement of the Universe*. Mackie writes: "In calling them eliminative methods Mill drew a rather forced analogy with the elimination of terms in an algebraic equation. But we can use this name in a different sense: all these methods work by eliminating rival candidates for the role of cause." (1980, 297)

his work on inductive logic (1980, Sec. 16 & 17). The latter but not the former has recently experienced some further developments (e.g. Kuipers 1984 and Romeijn 2006). While Carnap mostly treats very simple toy models, Keynes' attempts are closer to scientific practice and cover some real-world examples of case-to-case reasoning at least from a qualitative perspective. Several authors have suggested some general rules for analogical reasoning, for example[9]: (i) The more extensive the positive analogy, the greater the probability of the analogical inference. (ii) The more extensive the negative analogy, the smaller the probability of the analogical inference. (iii) The more extensive the implication, the smaller the probability of the analogical inference. While these rules certainly fall short of a full-blown framework, they do provide some systematic access to reasoning with analogies beyond mere heuristics.

To illustrate this, consider again the poor individual who died after drinking the herbal potion, which in its framing is a simplified version of the AIDS-example discussed in Section 3. Now, another individual drinks from a similar potion and the question arises whether he will die or not. The positive analogy consists in all those properties that both cases have in common, e.g. that the drink contained the same herbs, that it was prepared in a similar way, or that the two individuals were both healthy men under forty. The negative analogy concerns all those properties that are different in both cases, e.g. the potion may have been served at different temperatures and only one of the men may be a pharmacist, while the other is a philosopher. For this example, the general rules stated above are very intuitive. In particular, the probability that the second man also dies should not decrease, if the potions were served at the same temperature. By contrast, the probability should not increase, if the mixture of herbs were slightly different.[10] Finally, if the prediction is rendered more specific, e.g. that the death must occur within 48 hours, the probability of the inference will in general be smaller.

In many epistemological treatments of induction, an ill-conceived focus on enumerative induction prevails. The idea of a mere repetition of instances suggests that these are all exactly alike. Thus, enumerative induction precludes from the beginning inferences based on only partial similarity. Herein lies one of the principal reasons why the inclusion of analogy in a general framework of induction is believed to be so difficult. By contrast, the long neglected eliminative induction in the tradition of Mill's methods constitutes a natural approach for reasoning from similarity and dissimilarity. At least in some accounts, it allows for inferences in the presence of a negative analogy by introducing a rule for determining the causal irrelevance of circumstances (cp. Pietsch 2014, Sec. 3c). Now, if the known and unknown negative analogy only concerns properties that are irrelevant to q, then the analogical inference will be correct. By contrast, if at least one causally relevant property pertains to the negative analogy, then the analogical inference will fail.[11] In this way, eliminative induction provides a framework for treating predictive analogical inferences, while enumerative induction entirely fails to make sense of them.

[9] Cp. Keynes 1921, Ch. XIX; Bartha 2013, Sec. 3.1
[10] Note that this need not be the case for all properties. For example, if the second man were ill, the probability for his death would presumably increase. This underlines not so much the heuristic nature of analogical inferences but the need for a two-dimensional framework as outlined below.
[11] Additional complications may arise in the case of plural causation.

In principle, the sketched framework is deterministic, but probabilistic considerations can be integrated in a straight-forward manner: (i) when there are properties in the negative analogy of which one does not know whether they are causally relevant to q, as in the case of the temperature of the potion, which may for example destroy a poisonous substance; (ii) or when there are relevant properties in the unknown analogy, e.g. we may not know whether both men are healthy. In the first situation one needs to determine the probability that a property in the negative analogy is relevant or not, in the second situation the probability that a relevant property in the unknown analogy belongs to the negative or positive analogy. Of course, the combined situation can also occur of a property in the unknown analogy of which it is unknown if it is relevant. For determining these probabilities, the usual toolbox is available, in particular relative frequencies and symmetry considerations.

As a simple example, consider an analogical inference regarding two instances which differ only in irrelevant properties plus one property c of which it is not known whether it is causally relevant or not to phenomenon q. In one instance, q occurs in the presence of c and we are interested in the probability that in another instance, q will occur in the absence of c. Certainly, if we know the probability p that c is causally relevant for q, then the probability for the analogical inference, i.e. to observe q in the second instance, will be (1-p). For example, a specific ingredient may be missing from the potion in the second instance, but otherwise the situation shall be identical both in terms of the individuals drinking the potion and the way it is prepared and served. Now, if we can determine the probability that this additional ingredient is causally relevant, then we know how likely the death of the second person is, given the death of the first. Of course, substantial work is required how to determine and interpret these probabilities.

Or consider a situation, where all properties that are relevant to q belong to the positive analogy except for one c in the unknown analogy. Given the probability p that c is in the positive analogy, i.e. (1-p) that it is in the negative analogy, then the probability for the analogical inference of q will be p. Here, one option to determine p explicitly refers to the known positive and negative analogy, examining if there is any causal, statistical or deductive connection between properties in P and N with c. For example, all properties in P and N might be independent of c except one x in the positive analogy. Now, if it is known that x entails c with probability p, then we would have p for the analogical inference.

As an example, consider again the two men taking a sip from a herbal potion. The situations shall be alike in all relevant circumstances except that this time there is uncertainty, whether a certain crucial ingredient c of the potion is present in the second instance. The uncertainty determines how likely the second person is going to die. Some evidence may suggest the respective probability. For example, the color x of the potion may be known for both instances while being statistically correlated with the presence of the crucial ingredient c.

Thus, causal or deductive connections between properties constitute an important criterion to determine the probability for an analogical inference. That the internal causal structure between properties is important for analogical reasoning has been emphasized in particular by

Mary Hesse (1966, p. 59) and more recently by Paul Bartha (2010)[12]. A second criterion concerns the number of properties in the positive analogy compared with the negative and the unknown analogy, as is evident from the three rules that were given in the beginning of this section. Of course, there are enormous difficulties with adequately determining these ratios, e.g. detecting the independence of properties as well as their respective evidential weight. As also Norton (2011) emphasizes, the recognition of both criteria lies at the basis of the important two-dimensional model of analogical reasoning, where the horizontal relation determines the amount of similarity between instances in terms of properties and the vertical relation determines the nature of the connection between the properties.

In summary, eliminative induction provides a framework for reasoning by analogy or similarity in the predictive mode. Thus, the epistemic uncertainty connected with most analogical inferences stems not from the absence of a common logic but rather from the fact that the available evidence rarely is good enough for a quantitative assessment. The whole distinction between (enumerative) induction and analogy was ill-conceived from the beginning. Every induction is based at least partly on reasoning by similarity, because there are no instances that are alike in all respects. Eliminative induction can account for this.

The analysis of mechanisms is a further central element in the predictive mode of reasoning with case studies, where broadly understood mechanisms trace the relation between input and output quantities by looking at the detailed arrangement of parts and features. The study of mechanisms traces the causal processes in a case study and thereby further corroborates any causal relations between input and output quantities that may have been identified by eliminative induction. In the social sciences, the analysis of mechanisms is often referred to as process tracing, on which there exists abundant literature (e.g. George & Bennett 2005, Ch. 10). With respect to the two-dimensional model sketched above, mechanistic reasoning concerns the vertical relations, how the various properties within a case are connected.

In the philosophy of medicine, a current debate tries to understand the relation between comparative and mechanistic evidence (Russo & Williamson 2007). Very briefly, my viewpoint is that there is no real opposition, both can again be understood from the perspective of eliminative induction. Comparative reasoning establishes causal factors for a specific phenomenon, while mechanistic reasoning corroborates the relevance of these factors by analyzing more fine-grained connections. Well-established mechanisms provide additional confirmation for the causal relevance of factors by making the link to other bodies of evidence, i.e. essentially by unification. For example, explicating the chemical mechanism of a certain medication links up the causal effects of the medication with a well-confirmed system of fundamental chemical laws.

While the predictive mode of reasoning with case studies is ubiquitous in the natural sciences, it is not so prominent in the history of science. This is mainly due to the immensely complex causal structures with which historians deal that seldom allow the identification of sufficiently similar cases. However, when developing a case study, historians often rely on more or less

[12] "The validity of [an argument by analogy] will depend, first, on the extent of the positive analogy compared with the negative [...] and, second, on the relation between the new property and the properties already known to be parts of the positive or negative analogy, respectively." (Hesse cited in Norton 2011, p. 9)

predictive analogies with respect to other episodes in their background knowledge, to determine where to look for interesting material and which research questions to ask.

4c. Generalizing in the conceptual mode

As also shown by the role of exemplars in the natural sciences, case-based reasoning is often employed for concept development rather than for predictions. This observation is closely related to an interesting distinction made by some social scientists between eliminative and analytic induction (e.g. Hammersley et al. 2000). The basic idea in the latter is that one constantly needs to reformulate the hypothesis and redefine the phenomena when examining various instances in order to arrive at universal hypotheses – drawing attention to the often neglected fact that induction always goes hand in hand with conceptual refinement. Compared with eliminative induction, which mainly proceeds by parameter variation as discussed in the previous section, concept development is a much more complex and varied process. In fact, no general rules and principles seem to exist and presumably, it involves considerable creativity and intuition.[13] Historical studies suggest that the social and psychological context within which scientists work also plays a considerable role if only by providing a set of possible analogies.

Obviously, concept development can occur at different levels of coarse-graining. It can concern observational terms, but also more abstract generalizations. In the following, the focus lies mainly on the latter, henceforth referred to as a process of abstraction. The basic idea is that the conceptual scheme developed for one case can be transferred to another case – just as in the example of Maxwell's use of hydrodynamic analogies for developing electromagnetic field theory. With sufficient ingenuity, it is often possible to develop any conceptual analogy to a certain extent, but of course, not all frameworks are equally fruitful.

In abstraction, there are always different ways how to conceptualize, how to tell a story. In the natural sciences, this is known as the underdetermination of scientific theories. Raphael Scholl has recently pointed out an analogous predicament for &HPS illustrating his claims with the example of Semmelweis' discovery of the origin of childbed fever which can be reconstructed according to different methodologies, e.g. hypothetico-deductive or causal (Scholl 2013, Scholl forthcoming). Both in the sciences and in &HPS, the pluralism of interpretations is closely connected with the fact that in an abstractive mode some characteristics of a phenomenon are highlighted and others neglected depending on research interests. The resulting framework has to be evaluated in terms of pragmatic criteria, most importantly simplicity of description and fruitfulness for further research.

Besides underdetermination, a further difficulty in the conceptual mode concerns the theory-dependence of whether a case study is particularly representative or deviant. For example, the pendulum, the inclined plane, or the trajectory of a projectile are paradigmatic cases for classical mechanics, while being hard cases in Aristotelian physics with its concept of motion directed towards a natural place. By contrast, burning fire or falling rain were representative phenomena for Aristotelian dynamics but are largely uninteresting in classical mechanics. As

[13] For a good overview from a philosophy-of-science perspective consult Nersessian (2010), who stresses the role of analogy, imagery, and thought experiments. See also Carl Hempel's classic treatise on the subject (1952).

an example from &HPS, various methodological paradigms like hypothetico-deductivism or causal analysis each have their paradigmatic and hard cases.

It is beyond doubt that one can generalize conceptually from case studies. Rather, the crucial question is how general such a conceptual level should be. In retrospect, attempts at very high-level theories in history, such as the Marxist conception of historical law, have proved of dubious merit at best. They explain everything and nothing. A similar failure of high-level ideologies can be observed in the social sciences. Presumably, the main reason lies in the complexity and contextuality of the phenomena in both fields. By contrast, high-level conceptualizations are possible in physics, supposedly because it deals with much simpler phenomena.

Thus, a number of scholars have argued that historians should not aim at very general theories but rather at "mesoscopic" concepts that are well-adapted to specific contexts and can be subject to change over time as science evolves. For example, Peter Galison favors a study of "history claiming a scope intermediate between the macroscopic (universalizing) history that would make the cloud chamber illustrative of all instruments in all times and places and the microscopic (nominalistic) history that would make Wilson's cloud chamber no more than one instrument among the barnloads of objects that populated the Cavendish Laboratory during this century." (Galison 1997, 61)[14] Richard Burian makes a related suggestion: "Case studies cannot and should not be expected to yield universal methodologies or epistemologies. Rather, they yield local or, better, *regional* standards, and fallible ones at that." (Burian 2001, p. 400) In a similar vein, Hasok Chang has argued for replacing the term 'case study' by the notion of 'episode' for much the same reasons (2011, pp. 110-111). For Chang, an episode denotes the variation on a theme and not a mere instantiation of a general concept. Thus concepts have no static meaning, but have to be contextualized, while still allowing for inferences from one episode to the other.

Remarkably, a comparable suggestion was made by Robert Merton, arguing for the importance of middle-range theory in the social sciences: "Middle-range theory is principally used in sociology to guide empirical inquiry. It is intermediate to general theories of social systems which are too remote from particular classes of social behavior, organization, and change to account for what is observed and to those detailed orderly descriptions of particulars that are not generalized at all. Middle-range theory involves abstractions, of course, but they are close enough to observed data to be incorporated in propositions that permit empirical testing. Middle-range theories deal with delimited aspects of social phenomena, as is indicated by their labels." (Merton 1949, p. 531) Of course, coherence between different mesoscopic, regional, or middle-range concepts should still be sought. However, in view of the complexity of the phenomena both in history and the social sciences, one should abandon the search for extremely general high-level frameworks as in physics. Nevertheless, concept development remains an essential element of any scientific endeavor because trivially without concepts, one cannot account for anything.

[14] Galison argues for a "sited, not typical, history", the aim of which is "to evoke the mesoscopic periods of laboratory history, not a universal method of experimentation." (1997, p. 63) I largely agree but would add that scientific method nevertheless possesses a universal logical core, which has to be contextualized when analyzing specific episodes.

The problem of generalizing is thus very different for the conceptual and the predictive modes. One important aspect concerns the plurality of interpretations in the conceptual mode that is much less prominent when dealing with case-based predictions. Even though predictive statements may be phrased in different ways, their truth-value does not change. Another aspect concerns the criteria for evaluating analogical reasoning, which are of largely pragmatic nature in the conceptual mode, e.g. referring to simplicity and fruitfulness, while in the predictive mode empirical adequacy is the dominant factor. In a way, conceptualizations cannot turn out wrong, they just cease to be useful. While in the predictive mode, one is mainly interested in the reliability of inferences, the crucial question in the conceptual mode concerns the adequate level of generality such that the concepts are universal enough to be applicable in various contexts but not so general that they cease to be meaningful.

5. The vanishing boundary between case and statistical studies

The basic moral of the previous section is that detailed case studies are as important as cross-case comparisons for both reliable predictions and adequate concept development. It is mainly due to the limitations of the human cognitive apparatus that thus far one always had to make a choice between two complementary options how to deal with complexity: *either* by looking at a few cases in considerable detail *or* by looking at a large number of instances rather superficially. But in the end, the boundary between statistical reasoning and case studies is rather artificial. The best basis for both prediction and concept development would certainly involve a large number of cases in their full complexity.

Remarkably, there are some indications that the boundary between case studies and statistical reasoning is indeed beginning to dissolve due to the recent emergence of data-intensive science resulting from advances in information technology (Gray 2007; from a philosophy-of-science perspective e.g. Leonelli 2012, Pietsch Forthcoming). Although it is too early to really judge the impact of these developments, data-intensive methods are sometimes able to analyze a large number of cases in considerable detail.

An example concerns recent US election campaigns, in particular Obama's bid for office in 2012. Both Republicans and Democrats possess large data bases, in which for each citizen who is eligible to vote various data is gathered, e.g. on political preferences, demographic data and sometimes consumer data. In addition, some individuals volunteer information about how they voted in previous elections and on their current commitments. This data can then be used to make predictions about other individuals who have not disclosed their voting behavior. The algorithms providing for such predictions mostly rely on reasoning by analogy, they basically search for sufficiently similar individuals. Certainly, the characterization of voters in terms of a large number of parameters is still very crude. In this manner, it is hard to account for qualitative data or for the nature of connections between different properties. Still, data-intensive science currently constitutes the most promising approach to combine the analysis of within-case complexity with comparisons across a large number of cases.

At the moment, data-intensive science is still chiefly about predictions. Thus, its impact is felt especially in fields like the social sciences or medicine, where the predictive mode of reasoning with case studies is widely used. But I see no in-principle reasons why a data-based automation of concept development should not be possible. However, a much deeper theoretical understanding of the basic principles of concept development will be required. Also, much more complex knowledge architectures would need to be constructed than are currently available. After all, extensive background knowledge is required for the pragmatic evaluation of conceptual frameworks.

This means that natural scientists dealing with exemplars and philosophers of science reasoning on the basis of case studies will be fairly safe against any threat from information technologies taking over their work any time soon. However, the emergence of data-intensive science shows that it is high time to develop a common epistemological framework for case studies and statistical approaches as well as for analogy and induction.

6. Some final comments

Two modes of reasoning with case studies were sketched, one predictive, the other conceptual. While the latter provides the natural link between history and philosophy of science, the former plays only a minor role in &HPS, mainly in the day-to-day work of the historian, when deciding which sources to consult and what questions to ask by comparison with related cases in the background knowledge.

In the conceptual mode, case studies serve a number of purposes. They can ground methodological considerations in actual scientific practice. Historical case studies can provide a corrective to contemporary philosophical debates, they can probe and challenge philosophical theory. Case studies can also play a role in the discovery of novel theoretical phenomena and in forging new concepts to adequately account for these. A philosophy of science that cannot make a connection to scientific practice, both historical and contemporary, has no use whatsoever. Often philosophers have been much too quick in developing grand conceptual schemes on the basis of caricature-like toy examples instead of genuine history. On the other hand, a historical analysis that does not allow for methodological insights and generalizations is meaningless since there is nothing to learn from it. Thus, suggestions emerging from a direct confrontation of detailed history with larger philosophical claims, e.g. concerning the importance of mesoscopic theory, should be taken very seriously indeed.

Acknowledgments

I am much grateful to the editors of this volume, Tilman Sauer and Raphael Scholl, for helpful comments on the manuscript and also for organizing the inspiring workshop in Bern and contributing so many interesting ideas to the subject themselves. I also thank Christian Joas, Désirée Schauz, and Elsbeth Bösl for helpful discussions as well as Karin Zachmann for pointing me to the insightful discussion by Peter Galison.

References

Bacon, F. 1620/1994. *Novum Organum*. Chicago, IL: Open Court.

Bartha, P. 2010. *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. New York, NY: Oxford University Press.

Bartha, P. 2013. "Analogy and Analogical Reasoning." *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition). Available online: http://plato.stanford.edu/archives/fall2013/entries/reasoning-analogy/

Burian, R.M. 2001. "The Dilemma of Case Studies Resolved: The Virtues of Using Case Studies in the History and Philosophy of Science." *Perspectives on Science* 9(4):383-404.

Cartwright, N. 1983. *How the Laws of Physics Lie.* Oxford: Oxford University Press.

Carnap, R. 1980. "A Basic System of Inductive Logic." In R.C. Jeffrey (ed.): *Studies in Inductive Logic and Probability* (Berkeley, CA: University of California Press), 7–155.

Chang, H. 2004. *Inventing Temperature: Measurement and Scientific Progress.* Oxford: Oxford University Press.

Chang, H. 2011. "Beyond Case-Studies: History as Philosophy." In S. Mauskopf & T. Schmaltz (eds.): *Integrating History and Philosophy of Science* (Dordrecht: Springer), 109–124.

Duhem, P. 1954. *The Aim and Structure of Physical Theory.* Princeton, PA: Princeton University Press.

Galison, P. 1997. *Image and Logic: A Material Culture of Microphysics.* Chicago: University of Chicago Press.

George, A.L. & A. Bennett. 2005. *Case Studies and Theory Development in the Social Sciences.* Cambridge, MA: MIT Press.

Gerring, J. 2007. *Case Study Research. Principles and Practices.* Cambridge: Cambridge University Press.

Gray, J. 2007. "Jim Gray on eScience: A Transformed Scientific Method." In T. Hey, S. Tansley & K. Tolle (eds.): *The Fourth Paradigm. Data-Intensive Scientific Discovery* (Redmond, WA: Microsoft Research).

Hammersley, M., P. Foster, & R. Gomm. 2000. "Case Study and Theory". In R. Gomm, M. Hammersley & P. Foster (eds.): *Case Study Method: Key Issues, Key Texts* (London: Sage), 234–258.

Hempel, C.G. 1966. *Philosophy of Natural Science.* Upper Saddle River, NJ: Prentice Hall.

Hempel, C.G. 1952. *Fundamentals of Concept Formation in Empirical Science.* Chicago, IL: University of Chicago Press.

Hesse, M.B., 1964, "Analogy and Confirmation Theory," *Philosophy of Science* 31:319–327.

Hesse, M.B. 1966. *Models and Analogies in Science.* Notre Dame, IN: Notre Dame University Press.

Keynes, J.M. 1921. *A Treatise on Probability.* London: Macmillan.

Kuhn, T.S. 1977. *The Essential Tension.* Chicago, IL: Chicago University Press.

Kuhn, T.S. 1992. *The Copernican Revolution.* Cambridge, MA: Harvard University Press.

Kuhn, T.S., 1996, *The Structure of Scientific Revolutions, 3rd ed.*, Chicago: University of Chicago Press.

Kuipers, T.A.F. 1984. “Two Types of Inductive Analogy by Similarity.” *Erkenntnis* 21:63-87.
Leonelli, S. (ed.). 2012. Data-driven Research in the Biological and Biomedical Sciences. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43(1).
Mach, E. 1991. *Erkenntnis und Irrtum.* Darmstadt: Wissenschaftliche Buchgesellschaft.
Mackie, John L. 1980. *The Cement of the Universe.* Oxford: Clarendon Press.
Merton, R.K. 1949. “On Sociological Theories of the Middle Range.” In C. Calhoun, J. Gerteis, J. Moody, S. Pfaff, & I. Virk (eds.): *Classical Sociological Theory* (Oxford: Wiley-Blackwell 2012), 531-542.
Mill, J.S. 1886. *System of Logic Ratiocinative and Inductive*. London: Longmans, Green, and Co.
MMWR. 1981. “Pneumocystis Pneumonia – Los Angeles.” *Morbidity and Mortality Weekly Report* 30(21): 1-3.
Nersessian, N.J. 2002. “Maxwell and ‘the Method of Physical Analogy’: Model-Based Reasoning, Generic Abstraction, and Conceptual Change.” In D. Malament (ed.): *Reading Natural Philosophy* (Chicago, IL: Open Court).
Nersessian, N.J. 2010. *Creating Scientific Concepts.* Cambridge, MA: MIT Press.
Norton, J. 2011. “Analogy.” Draft chapter of a book on the material theory of induction. Available online: http://www.pitt.edu/~jdnorton/papers/material_theory/Analogy.pdf
Pietsch, W. 2014. “The Nature of Causal Evidence Based on Eliminative Induction.” *Topoi* 33(2):421-435.
Pietsch, W. Forthcoming. “Aspects of Theory-Ladenness in Data-Intensive Science”, *Philosophy of Science*. Preprint: http://philsci-archive.pitt.edu/10777/1/pietsch_data-intensive-science_psa.pdf
Pitt, J.C. 2001. “The Dilemma of Case Studies: Toward a Heraclitian Philosophy of Science.” *Perspectives on Science* 9(4):373-382.
Romeijn, J.W. 2006. “Analogical Predictions for Explicit Simmilarity.” *Erkenntnis* 64(2):253-280.
Russo, F. & J. Williamson. 2007. “Interpreting Causality in the Health Sciences.” *International Studies in the Philosophy of Science* 21(2):157-170.
Scholl, R. 2013. “Causal Inference, Mechanisms, and the Semmelweis Case.” *Studies in History and Philosophy of Science* 44(1):66–76.
Scholl, R. Forthcoming. “Inference to the Best Explanation in the Catch-22: How much autonomy for Mill’s method of difference?” Preprint: http://philsci-archive.pitt.edu/11041
Scholl, R. & T. Räz. 2013. “Modeling Causal Structures: Volterra’s Struggle and Darwin’s Success.” *European Journal for Philosophy of Science* 3(1):115–132.
Thomas, Gary. 2011. *How to do Your Case Study. A Guide for Students and Researchers.* Los Angeles: Sage.
Yin, R. K. 2009. *Case Study Research. Design and Methods.* Los Angeles: Sage