**Pluralism in Historiography: A Case Study of Case Studies[[1]](#footnote-1)**

**Katherina Kinzel**

[Final preprint. The final publication is available at Springer, ISBN 978-3-319-30229-4]

**Abstract**

In history the same historical episodes can be reconstructed from multiple perspectives, leading to different interpretations and evaluations of the same events, and sometimes even to different factual claims. In this paper, I analyze what I call “historiographic pluralism” – situation of conflict between different case studies of the same episodes. I address two interrelated questions: First, which features of historical reconstruction and representation give rise to such conflicts? Second, can we assess rival historical case studies and decide between them, thus restricting historiographical pluralism?

As an answer to the first question, I highlight the selective and theory-laden character of historical representation and argue that the narrative dimension of historiography is central for the knowledge that a historical case study can convey. I then go on to analyze how – in practice – disagreement about historical facts emerges. I discuss four case studies paired around two historical episodes and show that conflicts arise from the selective, theory-laden and narrative aspects of historical methodologies.

The second question I answer by discussing different criteria for assessing historical accounts. I note a dilemma in the evaluation of historical reconstructions. On the one hand, there exist neutral and almost universally accepted evaluation criteria. But these criteria are weak and cannot always decide between conflicting accounts of the same episodes. On the other hand, there are stronger methodological criteria. Alas, they are often not neutral with respect to the substantial theoretical issues at stake in situations of conflict between different historical reconstructions. I conclude that because of this dilemma, we have to accept some degree of pluralism in historiography.

**Introduction**

The story of the Scientific Revolution in the 17th and 18thcenturies has been told many times. It has been reconstructed in a discontinuous narrative by Alexandre Koyré who described it as a fundamental intellectual transformation triumphing in the mathematization of nature (Koyré 1957). It has been told as an origin story by Herbert Butterfield for whom it marks the advent of modernity (Butterfield 1949). Other authors have presented the story emphasizing continuities between modern scientific views and medieval and Renaissance knowledge practices. For instance, Alistair Crombie argued that experimental science had been practiced first by medieval natural philosophers (Crombie 1953). And Frances Yates stressed continuities between the Hermetic-Cabalist traditions of natural magic and scientific empiricism (Yates 1964). More recently, the prevalence of microhistory has led to a destabilization of big picture narratives, calling into question the very notion of the Scientific Revolution (Secord 1993). “There was no such thing as the Scientific Revolution*,* and this is a book about it“ (Shapin 1998, 1), Steven Shapin claims in the introduction to his reconstruction of 17th century science.

Perhaps the existence of many and conflicting accounts of the Scientific Revolution is not surprising. After all, what is at stake here is nothing less than the origin of our modern world view, the identity of the scientific enterprise, and the status of science in Western society and culture (Cunningham and Williams 1993; Lindberg 1990; Porter 1986). Ideologically contested issues like these are bound to provoke disagreement. The existence of plural narratives of the Scientific Revolution may simply reflect the changing ideological evaluations of science and its place in society.

However, we encounter pluralism in the historiography of science not only when it comes to large scale historical transformations of great political and ideological significance. Local historical cases that, at least at first sight, appear to be ideologically innocuous have met with the same fate: they have been reconstructed several times, from a variety of different viewpoints, and they have come to support different philosophical conclusions. There exist rival case studies of events as local as a specific measurement procedure, an experimental derivation, or an episode of theory-choice. For example, Milikan’s oil drop experiments which measured the charge of the electron have been reconstructed from competing sociological and rationalist perspectives (Barnes, Bloor, and Henry 1996, 18–45; Holton 1978; Franklin 1986, 140–162). There exist diverging accounts of the historical fate of Mendel’s experimentally derived laws of inheritance (for an overview see Sapp 1990). And the victory of Lavoisier’s oxygen-based over Priestley’s phlogiston-based theory has been interpreted in light of different philosophical accounts of scientific change and theory choice –pluralist, structural-realist, rationalist and sociological (Chang 2012; Ladyman 2011; Musgrave 1976; Kusch forthcoming). In some cases, the various reconstructions of the same historical events are compatible and complement each other. But in the examples mentioned above, the different case studies are in open conflict. They involve incompatible factual claims, give competing causal explanations, carry opposed epistemic evaluations, tell different narratives and reach rival philosophical conclusions.

This paper deals with the situations of conflict between different case studies of the same historical episodes. It addresses two questions: First, which features of historical reconstruction and representation give rise to such conflicts? Second, how can we assess rival historical case studies and restrict historiographical pluralism? Although I believe that my answers to these questions apply to historiography in general, I focus my discussion on case studies in the history of science and in history and philosophy of science (HPS).

In order to answer the first question, I discuss the interpretative and constructive dimension of historiographical methodology. My account highlights the selective and theory-laden character of historical representation and argues that the narrative dimension of historiography is central for the knowledge that a historical case study can convey. Based on this account, I analyze in detail four case studies paired around two historical episodes, and the methodological strategies employed in these case studies. I show that disagreement about historical facts emerges from the selective, theory-laden and narrative aspects of historical methodology.

 The second question I answer by discussing different criteria for assessing historical accounts. I note a dilemma in the evaluation of historical reconstructions. On the one hand, there exist neutral and almost universally accepted evaluation criteria. But these criteria are weak. They cannot always decide between conflicting reconstructions of the same historical episodes. On the other hand, there are stronger methodological criteria that constrain historiographical pluralism more drastically. Alas, these strong criteria are often not neutral with respect to the substantial theoretical issues at stake in situations of conflict between historical accounts. Because of this dilemma, I argue, we have to accept some degree of pluralism in historiography.[[2]](#footnote-2)

My paper has four parts. In the first part, I indicate on a general level which features of historical discourse give rise to pluralism. In the second part, I present a fine-grained account of disagreement in historiography by analyzing in detail four historical case studies. In the third part, I proceed to the problem of assessing rival case studies and discuss different historiographical evaluation criteria. In the fourth part, I apply these criteria to the case studies analyzed in the second part. The upshot of my discussion is that although historiographical pluralism is limited, it cannot be completely eradicated.

**1. Sources of Historiographical Pluralism**

What is it about the character of historical representation that enables substantially different retellings of the same events? How are competing historical reconstructions possible? Part of the answer is that historical discourse is a constructive and interpretative endeavor, and that historians can draw on a variety of different methodological strategies when reconstructing the past. In this section, I discuss, on a general and abstract level, three features of historical discourse that can help to understand why conflicting accounts of the same events are possible: (i) selectivity, (ii) theory-ladenness and (iii) narrativity. I claim that these are not merely contingent features of historical discourse. In fact, they seem necessary. I shall discuss each in turn.

*(i) Selectivity.* Like the Borgesian map of the Empire that has the size of the Empire, and coincides point-to-point with its territory, the complete historical account is an absurdity. All historical reconstructions are selective, and they are selective in three important ways. First, a historical account, unlike past reality, has a clearly marked beginning and an endpoint. From the infinite series of historical processes and events, the historical account selects an episode or case that is identified with a finite time-span. Second, once the time-span is determined, the historical account selects some events within that time-span and treats them as constitutive of the episode, while excluding others. Third, of the events included, some are highlighted, while others are relegated to a subordinate status.

Selectivity is a necessary feature of all representation. And recent discussions of representation in science have analyzed in great detail the selective choices that structure how scientific models represent their target domain. I will draw on these discussions, and in particular on “pragmatic accounts of scientific representation”, in order to illuminate the ways in which historical accounts are selective.

In pragmatic accounts, representation is extended from a two-term relation that refers the representation to its target, to a three-term or four-term relation that also involves the user of the representation and the user’s aims and purposes. For example, Ronald Giere states that representation always occurs in the context of use. “S uses X to represent W for purposes P” (Giere 2004, 743). Or, as Bas van Fraassen puts it, “[t]here are no representations except in the sense that some things are used, made, or taken, to represent some things as thus and so” (van Fraassen 2008, 23). Use encompasses a wide range of different factors: “the intention of the creator, the coding conventions extant in the community, the way in which an audience or viewer takes it, the ways in which the representing object is displayed, and so forth” (Ibid.). The context of use also determines the selective choices made in the representation. The ways in which representations are selective are hence not arbitrary but systematically dependent on the context-specific relations between user, representation and represented.

I want to suggest that historical accounts are similar to scientific models in that they select and highlight specific aspects of their target domain at the expense of others. And the selective choices of historiography are not arbitrary. They are structured according to the aims a historical account sets out to fulfill. Take Koyré’s account of the Scientific Revolution as an example. One of Koyré’s aims was to prove the relevance of non-testable metaphysical conceptions in theory creation, and to show that progress in science is driven by conceptual changes. As Rivka Feldhay points out, Koyré’s historical narrative is perspectivally constrained. It includes and highlights some aspects of the historical process at the expense of others. Koyré represents history “as a series of texts involved in networks of dialogues” (Feldhay 1994, 37). Focusing on the consistency of historical texts and the ideas embodied therein, he decides “to ignore the traces of production in the texts” (Ibid.). His narrative brings ideas and texts into hermeneutic focus, but it excludes the social and cultural relations of their productions. These inclusions and exclusions are clearly in line with his general aims. The actual selections made are aim-dependent.

Other examples can be generated with ease. If our aim is to explain historical change then the selection of discontinuous features of the historical episode under study, and the isolation of factors that may be seen to have caused or motivated these changes is the obvious strategy one should go by. If the aim is to create an experience of “historical otherness” then the features of the historical situation under study that stand out as absurd or unintelligible in light of current beliefs should be emphasized. If the aim is to study historical phenomena in their “longue dureé” then large-scale features of the historical period need to be prioritized. And so on.

 The selective character of historical representation and the aim-dependence of selective choices constitute important sources of pluralism. If there are many different historiographical aims, then there can be multiple ways of reconstructing the same historical episode that satisfy different aims and make different selective choices about which events to include and which to highlight. Different selections lead to diverging accounts of the same historical events.

*(ii) Theory-ladenness.* Historical accounts are not only selective in that they include certain historical events to the exclusion of others. On a more fundamental level, what it means to be a historical event and what it means to be a historical fact are outcomes of constructive processes. The facts of history are not simply found but have to be inferred from historical sources through complex inferential and interpretative processes. These processes, in turn, are structured by theoretical presuppositions. Hence, historical events and facts are theory-laden in the sense that they are partly constituted by theory.[[3]](#footnote-3)

Theoretical assumptions structure all stages of the inferential process from the sources to facts. To begin with, judgments regarding which types of sources are relevant for the reconstruction of a given episode, as well as judgments regarding which types of sources are reliable are made by reference to theoretical background knowledge. Then, after the sources have been chosen, these sources need to be interpreted in a consistent manner: the sources need to be related to each other; past events, actions and meanings need to be inferred from them; relations (possibly causal) between the derived events need to be identified; and the events need to be assigned a certain significance with respect to each other or with respect to the present. And each of these interpretative maneuvers relies on theory. Moreover, there is not only an upwards inferential and interpretative process that leads from the sources to the facts, but also a downwards process in the concept-dependent identification of historical events. Each historical event is an event only under a description, and hence its identification is only possible if the identity conditions for the event are specified: “[E]vents may be sliced thick or thin, a glance may be identified as an isolated event or as an instance in an event. What the unit-event is depends on the telling of it”(Roth 1988, 9).

There are very likely many other forms of theory-ladenness in historiography. But these considerations suffice to indicate how theory-ladenness can be a source of pluralism: different theoretical assumptions and different methodological commitments will have consequences for the selection of relevant sources, for how historical events are reconstructed from the sources, for how they are interpreted, explained and evaluated, for the individuation of historical events, etc. Since historical facts are theory-laden, disagreement is likely to emerge between historical accounts that reconstruct the past on the basis of different theoretical assumptions.

*(iii) Narrativity.* A third and final feature that leads to pluralism is the narrative character of historical discourse. The narratological tradition within the philosophy of history has long claimed that the peculiar form of historical representation – that which distinguishes historical reconstructions from other types of representational discourse – consists in the use of narrative. Rendering past events, states and processes intelligible requires that a story be told about them. By becoming entrenched in a story historical events achieve significance and meaning and which story is told is relevant for the knowledge a historical account can convey.[[4]](#footnote-4)

 Hayden White observed that in order to build a historical account, the series of historical events (the chronicle) has to be molded into a story that characterizes these events in terms of beginning, transitional phase and endpoint(White 1973, 5). And according to White, in this process, the choice of narrative form or mode of emplotment is crucial. This is because narratives “familiarize” historical events by relating them to already established plot structures. Historical accounts elicit understanding by referring the unknown back to already known themes and structures embodied in archetypal story types, such as comedy, tragedy or satire. They integrate historical events into stories that are firmly entrenched in the cultural repertoire of known narratives and themes: “The effect of such encodation is to familiarize the unfamiliar, and in general this is the way of historiography, whose ‘data’ are always immediately strange, not to say exotic”(White 1978, 49). Moreover, narratives convey information about the past by effecting closure. While the chronological series of historical processes and events is infinite, a narrative reaches an endpoint. And historical events are rendered intelligible when the story reaches its resolution, when the questions raised at the beginning have been answered and the reader’s expectations have been satisfied or disappointed(White 1980, 24–27). By achieving closure, a narrative constitutes a more or less coherent, meaningful whole.

Historical narratives, according to White, do not provide the unique truth about the past. “[A]lternative, mutually exclusive and yet equally plausible” (White 1978, 55) narrative emplotments of the same events can be constructed. This is just the situation that we have been describing: the repertoire of culturally preexisting genres and story types is vast, and one and the same historical episode can be rendered intelligible in manifold ways by drawing on different story types and modes of emplotment.[[5]](#footnote-5) About the same events, many different stories can be told.

The more general question that follows from the discussion of selectivity, theory-ladenness and narrativity in historical discourse is how strong pluralism is or needs to be. White himself remains somewhat ambivalent about the strength of pluralism. In some passages he grants that there are epistemic constraints on what types of narrative can be plausibly told about a specific historical episode (White 1978, 47-48, 59). In other passages, however, he suggests that there are not any limits on narrativization; “we are free to conceive ‘history’ as we please, just as we are free to make of it what we will” (White 1973, 433). Here it appears that White is not only a pluralist, but also an anything goes relativist for whom there are no epistemic constraints whatsoever that would restrict what narratives we can meaningfully and plausibly tell.

My own pluralist thesis is not that radical. In later sections, I will show that there are important epistemic restrictions on the range of permissible alternative historical reconstructions. But before doing so, I want to examine historiographical pluralism *in actu* by studying in detail the ways in which rival historical methodologies lead to different accounts of past events.

**2. The Structure of Disagreement: Four Case Studies**

Pluralism is most interesting, or most controversial, when two conditions obtain; namely (a) when there exist conflicting accounts of the same historical episodes, and (b) when it is not obvious which of the different reconstructions is the correct, adequate or most plausible one.

In the previous section I discussed how historiography can fulfill condition (a) and discerned three abstract features of historical discourse that lead to pluralism. This section explores how these abstract features are realized in actual historiographical methodologies. It deals with the concrete structure of disagreement in four historical case studies: Harry Collins and Allan Franklin’s rival accounts of the early gravitational radiation experiments and Alan Musgrave and Hasok Chang’s different reconstructions of how phlogiston theory was abandoned. Exploring how these case studies fulfill condition (a), I try to remain as neutral as possible between the competing accounts. The problem of evaluation and the question to which extent the rival accounts satisfy condition (b) I address in later sections.

 I structure my analysis of disagreement by introducing two levels on which differences between rival case studies can be observed; (I) the level of factual claims, and (II) the level of methodological strategies. I argue that differences arising on the level of factual claims can be traced to differences in methodological strategies. It is on this level of methodological strategies that the abstract features of historiography discussed above unfold their pluralist effects. As I will show in my analysis of the four case studies, the reconstruction of past episodes of science involves (i) selective choices, (ii) theory-laden reconstruction procedures and (iii) techniques of narrative emplotment. Differences in selection, theory-ladenness and narrativization give rise to conflicts about what exactly happened in the historical episode under study.

This is, I believe, a general point about historiography: factual claims in history are always the result of complex methodological processes and these methodological processes always involve selection, theory-ladenness and narrativization. However, it needs to be noted that the case studies I have chosen belong to the HPS context and that each of them comes with an explicit philosophical agenda. It may be argued that case studies in general history and in the history of science differ from those produced in HPS because they are not to the same degree committed to explicit philosophical doctrines. This is certainly correct. In general history, and in professional historiography of science, theory-ladenness does not typically take the form of explicit philosophical concepts being applied to the interpretation of the historical material. However, I believe this does not make case studies in general history and the historiography of science less theory-laden. It only means that the theoretical assumptions that structure the reconstruction of the past, and the ways in which they do so, are more subtle than they are in HPS.

*The Gravity Waves Episode*

I begin my analysis with the dispute over the interpretation of the high-flux gravity waves episode. Harry Collins’ reconstruction of the early attempts to measure gravitational radiation experimentally is central to the sociology of scientific knowledge canon, because it highlights the relevance of social factors in the closure of scientific controversies. Allan Franklin formulates his own account of the episode in direct opposition to Collins, arguing that social factors were not necessary for the closure of the debate. The historical events both authors can agree on can be summarized as follows: *In the late 1960s and early 1970s, Joseph Weber developed the first gravitational wave detectors and claimed to have acquired positive results. In the years to follow, other laboratories tried to replicate his results with slightly different experimental setups, but they did not manage to reproduce his observations. By the late 1970s, Weber’s claim to have observed high fluxes of gravitational radiation had lost all credibility and was rejected by the scientific community.* Despite their agreement on these points, Collins and Franklin accounts differ so radically that the two authors take the case of Weber to support two conflicting philosophical doctrines. How is this possible? There are both obvious and subtle differences between the rival reconstructions.

*(I) Factual claims.* Collins and Franklin disagree on how and why Weber’s claims were rejected. They make different factual claims about the historical episode. Collins claims that the available evidence and rational arguments underdetermined the decision against Weber, and that, eventually, social processes led to the rejection of his results. Franklin claims the exact opposite. According to him, rational deliberation was causally sufficient for a decision to be reached. Both authors claim that the historical material in fact supports their respective views of the relevant causal relations. According to Collins, the powerful rhetorical intervention of one of Weber’s critics (scientist Q) was decisive in tipping the scales to Weber’s disadvantage (Collins 1985, 93–95). Franklin, by contrast, thinks Q’s rhetorical attack played only a minor role and argues that it was the sheer quantity of negative evidence against Weber that eventually led scientists to discard his results (Franklin 1994, 468–69).

*(II) Methodological Strategies.* To understand how these different factual claims come about, we have to take a look at the different methodological strategies that the two authors apply. In the following, I will argue that Collins and Franklin select different aspects of the past to be represented, reconstruct events on the basis of different philosophical assumptions and tell different narratives. This accounts for the disagreement between them.

*(i) Selection.* When Collins reconstructs the scientific debates surrounding the experiments, he not only focuses on how the scientists who attempted to replicate Weber’s initial results responded to his claims and arguments. He also presents in great detail how they responded to each other. And he reveals that scientists were in severe disagreement as to how to interpret and explain their failures to replicate Weber’s findings. They found fault not only with Weber’s experimental setup, but also with each other’s experimental strategies (Collins 1985, 84-88, 90-92). Collins selects those aspects of the debate that indicate that there was profound disagreement in the scientific community.

 But these aspects are almost completely absent from Franklin’s reconstruction. Franklin presents the arguments in such a way that they fall into two opposed camps: Weber and his critics. The interrelations between Weber’s critics and their mutual criticisms are not taken into account. Franklin is quite clear that, for him, the situation is one of agreement rather than disagreement: “The fact that Weber’s critics might have disagreed about the force of particular arguments does not mean that they did not agree that Weber was wrong.” (Franklin 1994, 472) So while Collins devotes much attention to the various points of disagreement between Weber’s critics, Franklin selects and highlights points of agreement rather than disagreement.

 Another salient difference regarding the selection of historical events concerns the role of scientist Q. Collins places Q at a central point in the narrative: Frustrated by the scientific community’s hesitance to reject Weber’s results, which he took to be mistaken from the beginning, Q sets out to destroy the credibility of Weber and his observation claims in a series of polemical attacks. In Collins’ reconstruction, Q’s rhetorical intervention constitutes the social cause that tips the scales against Weber and leads to the closure of the debate (ibid., 93-95).

Franklin, in contrast, emphasizes the continuity of negative results that existed before and after Q’s intervention. He excludes Qs intervention from his account. According to him, it was the accumulation of negative results that led to the eventual rejection of Weber’s claims, not the rhetorical intervention of one scientist.

 Collins and Franklin’s different factual claims are a result of the selective choices they make. These choices structure which historical events and which aspects of the scientific debates under study are included and emphasized in their rival historical reconstructions.

*(ii) Theory-ladenness.* A basic difference between the two authors concerns their handling of the sources. Collins extends the realm of sources from the published record to also include interviews with the scientists involved, while Franklin puts the emphasis on the published material. These decisions are informed by theoretical assumptions which the authors themselves make explicit. Collins draws on interviews so as to avoid publication bias (Collins 1994, 498), while Franklin believes the published record to be more reliable than other sources (Franklin 1994, 465).

 More complex forms of theory-ladenness can be observed in how the two authors interpret the historical material. Collin describes the disagreement among Weber’s critics by referring to the concept of the experimenters’ regress. The experimenters’ regress typically occurs at the frontiers of enquiry when new phenomena are measured with new experimental apparatus. In these situations, ascriptions of when the apparatus is working properly hinge on whether it produces the wanted outcome, while, at the same time, what the correct outcome is becomes defined by reference to the quality of the experimental setup (Collins 1981, 34; Collins 1985, 84).

 Because he applies the theoretical concept of the experimenters’ regress to the historical material, he can interpret the situation not only in terms of disagreement, but also in terms of contingency and open-endedness. He claims that the historical process could have taken a different trajectory than it actually did.

Obviously the sheer weight of negative opinion was a factor, but given the tractability, as it were, of all the negative evidence, it did not *have* to add up so decisively. There was a way of assembling the evidence, noting the flaws in each grain, such that outright rejection of the high flux claim was not the necessary inference. (ibid., 91)

Franklin engages his reconstruction of the historical events with a philosophical agenda diametrically opposed to Collins’. He seeks to show that the resolution of the debate was not contingent, but the only rationally acceptable outcome. He applies the concept of robustness to the historical material to argue his point. When Collins concluded that the plurality of interpretative options marked the situation as open-ended and contingent, Franklin believes that the different arguments reinforced one another. The fact that a series of slightly different experimental setups failed to replicate Weber’s claims renders the negative results more robust; it strengthens the argument against Weber.

 Again, the differences in factual claims can be seen to rest on different theory-laden reconstruction procedures. Collins and Franklin select different sources and then interpret these sources by reference to different theoretical concepts. Collins applies the concepts of the experimenters’ regress and contingency, while Franklin draws on the concept of robustness.

*(iii) Narrative.* Finally, Collins and Franklin tell different stories of the historical events. The conflicting factual claims they make rest on different narrative emplotments of the episode. In order to establish this point, I propose a slightly unconventional practice of analysis: I will read the historical case studies as one would read a novel or a short story and apply some basic lessons from literary criticism. This will enable me to identify the narrative structure of the respective case studies and to show how different narrative structures carry different claims about the past.

Proceeding in this manner, Collins historical narrative is best described as an ironic tragedy. It resembles a tragedy because it tells the story of a social downfall and does so in discontinuous terms. According to the literary theorist Northrop Frye, a tragic plot is essentially a story of exclusion in which the hero is expelled or isolated from his society (Frye 1957, 35–43). The story of Weber, as Collins presents it, is such a story: Weber is excluded by the society to which he tries to belong. Moreover, the tragic plot is usually discontinuous, characterized by a radical break: before and after Oedipus finds out that he killed his father and married his mother, before and after Lady Macbeth dies and Macbeth finally realizes he has been tricked by the witches. In Collins’ narrative the discontinuity is Q’s intervention. By structuring events in terms of discontinuity the tragic plot reflects the demand for identifying a causally decisive turning point.

 What makes the tragedy an ironic one is that the hero is not causally responsible for his fate. As Frye explains, “the central idea of tragic irony is that whatever exceptional happens to the hero should be causally out of line with his character” (ibid., 41) The hero’s demise is not brought about by a tragic hamartia: “Irony isolates from the tragic situation the sense of arbitrariness, of the victim’s having been unlucky, selected at random or by lot, and no more deserving of what happens to him than anyone else would be” (ibid.). This sense of arbitrariness resonates with the interpretation of scientific closure that Collins defends: Weber’s downfall is not a result of him being in error. What happened to him was at least partly due to circumstance and the story could well have ended differently. His was a contingent downfall. The historical claims Collins defends are thus embodied in the narrative structure of the account.

 Not surprisingly then, Franklin’s narrative is fundamentally different from Collins’. It can be read as an adventure story. Adventure stories organize time in a strictly serial order. As Mikhail Bakhtin observes, the adventure story is constructed “as a series of tests of the main heroes, tests of their fidelity, valor, bravery, virtue, nobility, sanctity and so on” (Bakhtin 1986, 11). Typically, the hero emerges victorious from each test and the story closes with the hero’s exaltation. Franklin’s reconstruction resembles an adventure story in that it is strictly serial and accumulative. Moreover, the stages of the historical development are constructed as tests for Weber. Weber’s claims are confronted with a series of problems to which he is forced to respond. In each test-situation the arguments for and against Weber are reconstructed in dialogical situations that juxtapose Weber’s own considerations with those of his critics, so as to mimic an exchange of arguments. Weber, as the anti-hero of the story, emerges defeated from each test. Franklin’s account turns the exaltation of the hero into his demise; his is an inverted adventure story.

 As in the case of Collins, the plot structure carries historical claims. In Franklin’s narrative, we witness the evidence against Weber – negative results, problems, doubts, criticism, counter arguments and errors – piling up as the story progresses. The “overwhelming negative evidence” (Franklin 1994, 472) against Weber makes it inevitable for a rational scientific community to reject his claims. The accumulative structure of Franklin’s text thus carries his explanation of the events in terms of negative evidence rather than social processes.

*The Chemical Revolution*

I hope to have shown that differences in factual claims go back to differences in methodological strategies and that in these methodological strategies, the abstract features of historical representation discussed before unfold their pluralist effects. To further substantiate this claim, I turn to a second example, the Chemical Revolution. Although this historical episode is more complex than the gravity waves case, my discussion will be relatively brief and schematic. I discuss and compare two reconstructions of the episode, Alan Musgrave’s Lakatosian rational reconstruction, and Hasok Chang’s attempt to mobilize the episode as a case in point for normative scientific pluralism. Here is what both authors can agree to have taken place in the Chemical Revolution: *The heyday of phlogiston-based explanations of combustion and calcination occurred between 1700 and 1790. In the early 1770s however, Antoine Lavoisier began to develop an alternative explanatory framework that dispensed with phlogiston and instead postulated the existence of another substance, namely oxygen. Both the phlogiston and the oxygen theories enjoyed explanatory and predictive successes, as well as demonstrating appreciable problem-solving abilities. However, both frameworks also faced anomalies and failures. Precise weight measurements in later experiments favored the oxygen-based framework. Eventually, phlogiston theories were abandoned.* Beyond this basic agreement, Musgrave and Chang offer rival accounts of the processes through which phlogiston-based theories were replaced with oxygen based ones, rival explanations of why this occurred and competing epistemic evaluations of the rationality and legitimacy of the victory of oxygen.

*(I) Factual Claims.* Musgrave claims that phlogistonism constituted a degenerating research program and was rejected for that reason. According to him, the Chemical Revolution was a rational process (Musgrave 1976, 205–06). Chang, by contrast, claims that the phlogiston theory was not clearly inferior to its competitor (Chang 2012, 19–29), and that its potential had not been fully exhausted at the time of its abandonment (Ibid., 43–48). For these reasons, the rejection of phlogiston was a non-rational and premature decision.

*(II) Methodological Strategies.* As before, an analysis of methodological strategies will provide a better understanding of how such different factual claims can emerge.

*(i) Selection.* Musgrave and Chang delineate the episode of the Chemical Revolution in different ways. Two aspects of their selective choices are particularly salient. First, Chang situates the rejection of phlogiston theory within the broader context of a long-term transformation of epistemic practices, the rise of what he calls the compositionist system of Chemistry. According to Chang, the rejection of phlogiston theory was not rational in itself, but a mere epiphenomenon of this broader shift (ibid., 36-42). Musgrave, in contrast, chooses to represent the concrete interactions between Priestley, Cavendish and Lavoisier, but he excludes long-term transformations in Chemists’ practices from his account.

 Second, Chang chooses to represent not only the actual historical events, but also what could have become of phlogiston theory, had it not been abandoned. He presents a counterfactual history that phlogiston theory could have fostered scientific progress had it been retained. In this ways, the counterfactual history becomes part of the episode under study.

 It is by including the broader context of epistemic transformations in Chemistry and the counterfactual benefits of phlogiston that Chang can claim the rejection of phlogiston theory to have been non-rational and premature. Musgrave and Chang’s different verdicts on the rationality of the Chemical Revolution are underpinned by the different selective choices they make.

*(ii) Theory-ladenness.* A fundamental difference between Musgrave and Chang’s reconstructions concerns temporality. Musgrave reconstructs the development of the phlogistonist and oxygenist rivals diachronically, distinguishing between different successor versions of the theories. He applies Lakatos’ conception of competing research programs to the historical material; and a Lakatosian research program consists in a diachronic series of successor versions of a theory.

 Musgrave also uses the concept of progressive and degenerating research programs to evaluate the rivals. In his interpretation, both programs were successful before 1770, and it was only between 1770 and 1785 that Lavoisier’s oxygen theory began to outperform the phlogistonist program of Priestley and Cavendish. During that time span, the oxygen program developed in a coherent manner, according to Musgrave, each new version marked an increase in predictive power and theoretical growth. The program was progressive. The phlogiston program, in contrast, was confronted with increasing difficulties and degenerated (Musgrave 1976, 205). This evaluation then allows for the verdict that the rejection of the phlogistonist explanations was entirely rational: chemists at the time realized that the phlogiston-based system was degenerating and changed their allegiances.

Chang reaches a very different verdict. This is possible primarily because his reconstruction is systematic rather than temporal. Chang does not recount the successive steps in which the two theories developed. He applies a different concept to the historical material, reconstructing the rivals as holistically understood systems of practice. He analyses static and systematic features of the phlogistonist and oxygenist approaches, listing the questions the two systems addressed, the problems they found significant and the epistemic values they embodied (Chang 2012, 19–28).

Comparing the two systems, Chang applies the concept of methodological incommensurability. According to his interpretation, both systems were able to solve the problems which they considered important in a manner that satisfied the epistemic values that they adhered to.

It seems clear that each of the oxygenist and phlogistonist systems had its own merits and difficulties, and that there were different standards according to which one or the other was better supported by empirical evidence. […] [B]oth systems were partially successful in their attempts to attain worthwhile goals and […] there was no reason to clearly favor one over the other. (Ibid., 29)

As in the above example, we can observe how differences in the theoretical assumptions that guide the historical interpretation translate into different factual claims: Musgrave engages in a diachronic reconstruction and interprets the situation in terms of Lakatosian confirmation theory, using the concept of research programs. On this basis he reaches the conclusion that the Chemical Revolution was a rational process. Chang, in contrast, interprets the situation in terms of the theoretical concept of methodological incommensurability and finds that the decision was not rational.

*(iii) Narrative.* As in the example above, the historical claims are also brought across by the way the story is told. Musgrave’s historical account is well described as a comedy. In a comedy, according to Frye, the complications, plot twists and revelations that the hero has to live through before succeeding are more important in determining the course of events than the moral or intellectual qualities of the characters(Frye 1957, 170). Just like a comedy, Musgrave’s narrative is driven not by the intentional action of the characters, but by unexpected plot twists. From the moment that Lavoisier’s research program enters the stage, we are confronted with a series of sudden and unexpected changes of fate, some of which reveal a deep historical irony. Repeatedly, Lavoisier’s opponents are also his helpers. They provide the insights which enable Lavoisier to verify his prediction (of what would later be called oxygen), and later, to solve the anomalies that trouble his oxygen program (Musgrave 1976, 194, 200–01). The comic emplotment in terms of unexpected plot twists accords with Musgrave’s interpretation of the episode as rational. Scientific rationality exerts itself in the narrative as a Hegelian “List der Vernunft” (cunning of reason). Even when trying to disprove the oxygen theory, its opponents only contributed to its victory. And despite occasional contingencies and surprising plot twists, the more successful program wins eventually.

Chang’s historical account draws on completely different narrative principles. The story he tells is not comic; to the contrary, the structure of his text is best captured if it is read as an elegy – an elegy to a promising theory that died an unjust death. Indications of such a reading can be found in Chang’s own headlines, which read “The Premature Death of Phlogiston” (Chang 2012, 1) and “Why Phlogiston Should Have Lived” (ibid., 14). While life and death are no more than metaphors that facilitate easy comprehension of the main judgments that the historical argument is supposed to put forward, they are instructive. Like an elegy to a prematurely deceased hero, Chang’s narrative mourns unrealized possibilities. It conjures up an image of what could have become of the deceased had they not passed away and seeks to establish that phlogiston, in line with Chang’s normative pluralist agenda, should not have been abandoned. To prove this point, Chang engages in counterfactual history about what might have happened had phlogiston theory been maintained (ibid., 42-50). This counterfactual imagination enables the judgment that the death of phlogiston was not only not rationally warranted (and in this sense unjust) but also premature if measured against the possibilities for innovation, development and discovery that it entailed. The elegiac practices of bemoaning and praising the deceased fit Chang’s evaluation of the situation like a glove.

Having provided a detailed analysis of selected historical case studies, I hope to have made plausible two points. The first point is that there is not one unproblematic way of deriving historical facts from the sources. Rather, historians engage in complex methodological strategies in order to reconstruct, interpret, evaluate and explain past events, and the methodological strategies of historiography involve selective choices, theory-laden interpretations and narrative emplotments.

 Second, I hope to have shown that differences arising on the level of factual claims can be traced to differences in methodological strategies. Which selections are made when reconstructing historical happenings, which theoretical assumptions guide the interpretation and evaluation of past events, which narratives structure the historical material has consequences for what factual claims a historical account can reach. Given these features of historiography, the existence of severe disagreement between case studies of the same episodes does not appear surprising.

**3. Evaluating Historical Accounts**

In my analysis above, I tried to stay as neutral as possible between the rival historical accounts, and did not present one side of the conflict as more plausible or better warranted than the other. In this section, I turn to the problem of whether and how we can assess historical case studies and settle conflicts such as those discussed.

As stated before, historiographical pluralism is most interesting if (a) it occurs between conflicting historical accounts, and (b) the alternatives on offer are plausible to roughly similar extents. The question as to whether and how we can assess competing historical case studies is of central importance to historiographical pluralism, because it bears on condition (b), the issue of comparative plausibility: the stronger our grounds for assessing case studies and for deciding between rival reconstructions, the weaker is our pluralist scenario. If we can always reach unequivocal decisions between competing accounts, then there is no room for pluralism in historiography. Or at least the more controversial forms of pluralism that occur between incompatible and conflicting reconstructions will be ruled out. On the other extreme, if we can never judge which of two or more alternative accounts is the most plausible, then we are confronted with a situation of extreme pluralism, or even anything-goes relativism.

My own approach takes a middle position between these two extremes. I argue that there indeed are epistemic considerations that allow for an evaluation of competing historical case studies. These considerations place restrictions on the space of permissible alternatives and hence restrict pluralism. However, they are not strong enough to always yield an unequivocal verdict as to which of two or more competing reconstructions to prefer. In some cases, a neutral decision between rival accounts may not be possible. In order to make my point, I begin by considering what types of epistemic considerations we can draw on in order to evaluate historical case studies. I distinguish between (α) basic and (β) complex evaluation criteria. Then I proceed to evaluating the above discussed case studies in terms of these considerations.

*(α) Basic criteria.* When historians and philosophers of science discuss the merits of different historical reconstructions and case studies, their evaluation criteria often remain implicit.[[6]](#footnote-6) However, there seem to exist a few rough and ready rules that one can draw on when assessing the quality of a historical reconstruction. Some of them are related to the practices of source criticism (how reliable are the sources used, how well are the known sources covered, and how varied is the evidence cited?). Others concern the composition of the historiographical text itself (is the historical argument consistent and intelligible?). Yet others relate to how well the historical reconstruction fits within a broader system of knowledge (is the historical account consistent with accepted, incontrovertible background knowledge?). Standards regarding source-reliability and source-variance, internal consistency, and consistency with accepted background knowledge are relatively uncontroversial and they are commonly relied upon even when they are not made explicit. I refer to this type of evaluative standards as basic criteria.

The basic criteria have the advantage of being relatively neutral with regards to philosophical conflicts. That is, when using them to decide between conflicting historical accounts, we can usually be relatively certain that we are not already assuming a point at stake in the debate. Requirements for internal consistency merely raise demands concerning the logical or argumentative structure of the historical reconstruction. Such demands seem to be neutral in relation to the theoretical assumptions that might be at stake in a conflict between rival historical reconstructions.

Considerations regarding the reliability, variance and completeness of the sources are, as described above, not without theoretical presuppositions. Whether a historical account has covered the relevant sources to a sufficient degree and hence is “complete” depends upon existing background knowledge about what sources exist, which in turn depends upon other available historical reconstructions, archival research, and so on. Moreover, what counts as a relevant source is contextually determined since it depends upon the aims and purposes of the historical account that draws on these sources. Selectivity and the restriction of covered sources is legitimate in principle, if it accords with and is conducive to the aims of the historical account. Completeness only refers to the completeness of the sources relevant to the satisfaction of a specific historiographical aim.

Nevertheless, a historical account that involves more varied sources can be considered more robust than an account that restricts its sources to a specific type. Also, there may be clear violations of the contextually understood completeness requirement. For example, if known sources that would be relevant to the historiographical aim but which are not in line with the argument that the historical account wishes to carry along are excluded, then the selective choices may be considered biased. This would strongly undermine the plausibility of the historical account in question. I believe that although they are context-dependent and theory-laden, evidential considerations such as the ones just presented can sometimes serve as neutral evaluation criteria. At a later point I provide an example for what an evaluation in terms of contextually determined source completeness can look like.

*(β)* *Complex criteria.* But there are also more complex considerations that can be and often are used in the assessment of historical reconstructions. Complex criteria are evaluation standards drawn from debates about intricate historiographical issues such as contextualism, internalism, and externalism (are historical events adequately contextualized?); hermeneutics, understanding, and translation (are the historical actors’ conceptions and understandings faithfully reconstructed and appropriately conveyed?); textual interpretation (is the original meaning of the text restored?); present-centeredness (are backwards-projections, anachronisms and Whig-history avoided?); historical explanation (have the right causes been identified, has reductionism been avoided?); micro- and macrohistory (does the account address the correct level of description?) etc. Complex criteria, unlike basic ones, have the advantage that they are subject to explicit discussion. They are therefore relatively well understood and usually rendered explicit when they are used in the evaluation and critical assessment of a historical reconstruction.

On the downside, unlike basic ones, complex criteria are not generally agreed upon or uncontroversial. And disagreement regarding complex criteria can arise on at least two levels. First, it may not always be evident whether a complex criterion has been met. For example, it is not always evident whether illegitimate backwards projections and anachronisms have been avoided in a given reconstruction, or whether a historical account exhibits explanatory power. But second, and more importantly, the complex criteria themselves are contested.

For example, in discussions concerning Whig history and present-centeredness there is substantial disagreement regarding the identification of the exact vices that result from present-centered historiography of science. There is also disagreement concerning whether all or only some specific uses of present-day knowledge and categories in the interpretation and explanation of past science are to be avoided (Ashplant and Wilson 1988; Cunningham 1988; Cunningham and Williams 1993; Jardine 2000). Moreover, it has been suggested that the evaluation of past knowledge by present-day standards might not always be problematic (Tosh 2003), or at least that it is not as problematic as other practices that have come to be criticized under the heading of present-centeredness, for example, a triumphalist siding with the winners of past scientific debates (Chang 2009).

Regarding adequate levels of analysis and explanatory power, it is debated whether the capacity of the historiography of science to provide comprehensible explanations depends on micro-perspectives. Does historical explanation need to trace the local and particular causes that prompt specific historical events or can it be concerned with large-scale factors and processes as well? (For a useful discussion of the respective epistemic capacities of macro- and microhistorical perspectives see Pomata 1998). More fundamentally, it is not even universally agreed upon that explanation, and in particular causal explanation, should be a central aim and method in the historiography of science (For example, the relevance of causal explanations in history has recently been denied by Daston and Galison 2007, 34–37; for critical discussion see Kinzel 2012).

 Apart from not being generally agreed upon, the complex methodological criteria often are connected with theoretical assumptions about the character of science, or about the relations between past and present-day knowledge. They are connected with substantial philosophical issues and hence are seldom neutral with regards to philosophical debates.

 The case is most obvious with methodological debates concerning adequate contextualization. Clearly, in the dispute between Collins and Franklin, questions regarding the method of adequate contextualization are intimately related to what is ultimately at stake in the conflict between them: the social nature of scientific decision-making. The other complex criteria are philosophically laden in a similar manner. Methodological debates over anachronism and Whig-history are related to philosophical questions regarding continuity and discontinuity in the history of science, scientific change and progress. Considerations regarding explanation and understanding in history carry over into philosophical issues concerning the relations between reasons and causes, scientific rationality, and the driving forces of theory change. And to the degree that the complex criteria are not independent of philosophical positions and claims, they should also not be expected to be neutral with regards to the theoretical issues at stake in conflicts between rival historical reconstructions. When we are relying on a complex criterion in order to decide between conflicting historical accounts, we cannot always be sure that we are not already assuming a point at stake in the debate. In some cases, the failure of a historical account to satisfy a certain complex criterion may be more indicative of that same criterion being defined in a philosophically invested manner, rather than of a neutrally assessable flaw of the account in question.

**4. Constraints on Pluralism**

Having distinguished between basic and complex criteria, I want to return to the four historical case studies analyzed earlier in order to evaluate how well they fare with respect to the neutral evaluation criteria we introduced. In this analysis, I seek to substantiate two claims. First, basic evaluation criteria reduce the space of permissible alternatives and hence restrict pluralism. But second, the verdicts that we can reach on their basis are relatively weak, and in order to reach a more definite decision, we would have to refer to complex criteria.

I begin with applying the basic evaluation criterion of internal consistency to the competing historical accounts of the Chemical Revolution. The criterion of internal consistency restricts the space of permissible alternatives, because, of the two case studies I have discussed, only one meets its standard. Musgrave’s account fails the consistency requirement. It involves a straightforward contradiction in its central factual claims. This contradiction emerges as follows: At the end of the historical narrative, and after having given his presentation of the historical development of the two competing research programs, Musgrave passes the following verdict on them:

Between 1770 and 1785 the oxygen programme clearly demonstrated its superiority to phlogistonism: it developed coherently and each new version was theoretically and empirically progressive, whereas after 1770 the phlogiston programme did neither.(Musgrave 1976, 205)

This verdict is indispensable for assessing the abandonment of phlogiston-based theories as rationally warranted. And yet, this verdict does not accord with claims made earlier in the historical reconstruction. In particular, Musgrave had claimed that in 1775 experiments spoke as much against Lavoisier as they did in his favor (ibid., 196), and that in 1783 Priestley was having great predictive success with his phlogiston theory (ibid., 199). If these claims are correct, Musgrave’s statement that after 1770 the phlogiston theory was starting to degenerate while the oxygen theory was progressing cannot be right. The historical facts he cites contradict his interpretation of the situation in terms of progressing and degenerating research programs. The claim that the choice for oxygen was rational by Lakatosian standards appears ill-grounded. Chang’s account is less problematic in this respect. At least, it does not involve historical claims which directly contradict each other and hence it passes the basic criterion of internal consistency. Applying the basic criterion of internal consistency to the conflicting reconstructions leads to the exclusion of Musgrave’s case study and hence restricts historiographical pluralism.

In the dispute between Collins and Franklin, evidential considerations become crucial. As mentioned above, the two authors draw on different types of sources. Collins goes beyond the published sources to also include extensive interviews with the scientists involved in the episode under consideration. In fact, in the presentation of his historical account interview excerpts are much more prominent than published material, since the interpretations and arguments surrounding the gravitational radiation experiments are primarily reconstructed on this basis. The consideration behind this strategy is that of circumventing publication bias and gaining insight into the reasoning processes of scientists before they are straightened to fit the format of a scientific journal. Collins explicitly criticizes Franklin for only drawing on the published record. According to Collins, the restriction to scientific publications only is simply bad historiographical practice (Collins 1994, 497–99). Is Franklin’s decision to cover only the published sources indeed as problematic as Collins’ suggests? Does Franklin fail the basic standards of completeness and variance of the evidence?

I think he does, although the situation is complex because of the theory-laden character of source selection. The main problem with Franklin’s account is that it excludes a whole class of known sources which would in principle be relevant to the historical argument at stake. The restriction on published sources cannot be justified on the basis that they were the only ones available. On the contrary, the original account that Franklin wishes to disprove includes unpublished material. This means that Franklin draws on a subset of the types of sources used in the original account. How a less complete consideration of existing sources could be better suited for representing the actual process of scientific decision-making remains unclear and we should at least be skeptical whether Franklin’s account passes the completeness requirement. However, Franklin justifies his restriction on source material by contending that the arguments that scientists themselves find to be the most convincing, the strongest reasons that they had for accepting or rejecting certain findings, are most likely to be found in the publications. According to Franklin, the publications display those reasons that scientists “wished to have made as part of the permanent record” (Franklin 1994, 465). The published record thus serves as a filter for scientists’ actual beliefs. Franklin thus mobilizes the criterion of source reliability against the criterion of source completeness. He justifies his practice of excluding certain sources on the basis that they are less reliable than the ones he brings into focus.

The question is whether this stance is legitimate in light of Franklin’s own representational aim, namely that of reconstructing the rational process through which scientists arrived at their verdicts about Weber’s observation claim. To regard the restriction as legitimate in light of this aim, one would have to subscribe to the thesis that published results are the best indicators of the rational reasoning processes that brought them about, and that they are better equipped for this task than other types of sources, such as interviews, unpublished manuscripts, letters and notebooks.

This assumption is almost universally rejected in the historiography of science. But perhaps more worryingly for Franklin, he himself does not consistently uphold the methodological principle of primarily drawing on the published record. For example, in his case study of Millikan’s oil drop experiments Franklin engages in great detail with Millikan’s notebooks in order to interpret the former’s experimental judgments as rationally justified (Franklin 1986, 140–157). This makes the choice of sources for his reconstruction of the early gravitational waves episode seem arbitrary, rather than methodologically justified. Arguably, Collins’ account fares better in this respect. One may point out though that, although Collins’ sources are more varied than Franklin’s, they are still not robust enough, since the strong emphasis on interviews is not sufficiently balanced with other types of source material. In the conflict between Franklin and Collins’, neutral evaluation criteria favor Collins’ reconstruction, but they do so only by a thin margin.

This is where I turn to my second claim. Neutral criteria can decide some historiographical conflicts, but the decisions they yield are relatively weak. A stronger decision could only be reached by drawing on some of the complex evaluation criteria.

There are two reasons why basic criteria are weak arbiters. First, they are only necessary but not sufficient for a historical account to be plausible. Therefore they act only as constraints on the space of alternatives, but do not single out one account as the correct or most plausible one. To illustrate this point, let us return to the dispute between Musgrave and Chang. The application of the internal consistency criterion excludes Musgrave’s account from the range of legitimate historical reconstructions. But of course, this does not turn Chang’s account into the one unequivocally correct representation of what happened in the Chemical Revolution. On the one hand, internal consistency is not the only evaluation criterion we can draw upon and there may be many reasons to be critical of Chang’s reconstruction that have nothing to do with whether it is internally consistent or not (for two recent criticisms of Chang’s account see Blumenthal 2013; Kusch forthcoming). On the other hand, I have only considered two of the many different and possibly conflicting reconstructions of the Chemical Revolution. There exist myriad alternative retellings of that episode (a comprehensive overview of the past 50 years of historical writing about the Chemical Revolution can be found in McEvoy 2010), and even without having analyzed them in detail, I contend that at least some of them will meet the basic criteria. These criteria restrict the space of possible alternatives, but they do not shrink it down so radically that it would only include one permissible account. Even after having applied them, there is still room for historiographical pluralism.

The second reason why the basic criteria are weak is that they serve to evaluate only specific case studies, but not more general principles of reconstruction, interpretation and narrative emplotment. While Musgrave’s case study has been rejected on grounds of inconsistency, the Musgrave-type of historical analysis has not. Internal consistency is not endemic to a specific type of historical analysis. Could one not tell the story of the Chemical Revolution in a similar manner as Musgrave does, by reconstructing the diachronic development of phlogistonism and oxygenism as competing research programs and by emplotting historical events in a comic form, but without repeating his mistake? We may not be able to express with Musgrave’s vigor the conclusion that after 1770 the oxygen-based program was clearly superior. But we could probably still tell the story of the success of the oxygen program as one in which scientific rationality prevailed through complex plot twists.

Or think about the debate between Collins and Franklin. We have seen that Franklin’s reconstruction fails because the evidence he adduces is insufficient. However, when I analyzed the methodological differences between the two accounts, I argued that the most important differences do not concern the mere choices of sources, but rather, how historical facts are derived from these sources. I showed that Collins arranges and interprets his sources in such a way as to highlight disagreement and open-endedness, whereas Franklin reconstructs from his sources a historical situation of agreement and robust negative evidence. The fact that the set of covered sources is not exactly coextensive in Collins’ and Franklin’s reconstructions very likely facilitates them reaching diverging reconstructions and interpretations. However, is it not at least possible that such diverging reconstructions and interpretations could be reached even if the same sources were used? Maybe Franklin would have served his point better had he chosen the same sources as Collins, yet interpreted them according to his own methodological principles. The same seems to be true for narrative structure. Whether the story of the early searches for gravitational radiation is an inverted adventure story in which negative evidence piles up or whether it is a discontinuous tragedy does not seem to be completely determined by the available sources. We need to at least consider the possibility that Franklin could have told his adventure story on the basis of the same sources that were also used by Collins. Historians do enjoy some degree of freedom when it comes to choosing their methodological strategies and forms of narrative emplotment. The basic evaluation criteria help to identify the flaws in specific historical reconstructions. But they are not strong enough to rule out, on a more general level, specific methods of interpretation or forms of emplotment as clearly illegitimate.

In comparison, the complex criteria are significantly stronger as arbiters. First, considerations regarding methodological principles, such as contextualization, explanation and present-centeredness can restrict the space of permissible alternatives much more radically than basic criteria do. The set of historical accounts of the same episode that are internally consistent and that handle the known evidence in a satisfying manner will arguably be much larger than the set of historical accounts that, in addition to satisfying the basic criteria, are also appropriately contextualized (according to a specific understanding of relevant contexts), exhibit explanatory power (according to a specific criterion of explanatory power), avoid anachronisms (according to a specific definition of anachronism), etc.

Second, if we restrict the range of permissible methodological principles, we have ipso facto restricted the range of permissible types of historical reconstruction, not just the set of actually existing acceptable case studies. If we can show that the fault with Franklin’s reconstruction goes beyond his handling of the sources, and that it lies in how he uses present-day knowledge in the interpretation and evaluation of Weber’s arguments, then we have not only excluded Franklin’s particular historical case study, but any historical reconstruction that draws on similar reconstructive and interpretative principles. If we can show that there is something wrong with the practice of rational reconstruction in Musgrave’s account, then we have not only excluded this particular case study, but any account of the Chemical Revolution that draws on a Lakatosian conception of scientific rationality.

Hence, the complex methodological evaluation criteria are significantly stronger than the basic ones. But applying them brings us into the center of substantial philosophical conflicts about the nature of science, the relation between past and present knowledge, scientific rationality, theory change and progress. The complex criteria are strong, but they are highly controversial and anything but neutral with regards to philosophical conflicts.

**Conclusion**

There appears to be a dilemma when it comes to evaluating historical reconstructions. On the one hand, there are basic evaluation criteria such as source-reliability, range of the evidence cited and internal consistency. These criteria are relatively neutral with regards to higher-level theoretical and philosophical conflicts. However, these neutral evaluation criteria are not very strong. They restrict pluralism but only to a relatively low degree. On the other hand, there are complex evaluation criteria that are stronger than the basic ones and restrict pluralism more radically. But the complex criteria are themselves contested and are seldom neutral with regards to the fundamental issues that are at stake in a conflict between different historical reconstructions. Put in a nutshell: neutral criteria are weak, strong criteria are not neutral.

This implies that we will have to live with some degree of pluralism in historiography, at least if we wish our decisions between competing historical accounts to be grounded in neutral criteria that are shared by everyone who participates in the debate. This pluralism will be limited because there are at least some neutral considerations that can serve to exclude unacceptable historical accounts. However, even after the clearly illegitimate historical reconstructions that do not meet the basic criteria have been excluded, there can still be plural historical reconstructions of the same historical episodes that support different philosophical doctrines.

Of course, we may not wish to remain neutral in our evaluations of historical reconstructions. A convinced social constructivist may well find dubious the methodological principles that Franklin employs, the interpretations he reaches, as well as his narrative strategies. The constructivist may wish to reject Franklin’s historical account on the basis that it is internalist and present-centered and hence fails those complex methodological criteria that call for a more thorough contextualization and historicization of scientific debates. But in doing so, the constructivist has assumed some of the points at issue, namely that a reconstruction of the debate in terms of its technical contents only is deficient and that past beliefs should not be evaluated by present-day standards. The constructivist may have good reasons for holding these views, but a decision between conflicting case studies that is based on them is not a neutral decision. Mobilizing complex criteria in conflicts between historical accounts reinforces historiographical pluralism rather than eradicating it.

Acknowledgements

For suggestions and comments on earlier drafts of this paper I am grateful to Martin Kusch, Veli Mitova, Martha Rössler, Katharina Bernhard and Martin Strauss. I also want to thank the participants of the workshop “The Philosophy of Historical Case Studies” in Bern for their engaging discussions and criticism. Reasearch leading up to this paper was made possible by grants from the Austrian Science Foundation (FWF) (Project no.: P25069-G18) and the European research Council (ERC) (Grant no.: 339382).

**References**

Ashplant, T. G., and Adrian Wilson. 1988. “Present-Centred History and the Problem of Historical Knowledge.” *The Historical Journal* 31 (2): 253–74.

Bakhtin, Mikhail. 1986. “The Bildungsroman and Its Significance in the History of Realism (Toward a Historical Typology of the Novel).” In *Speech Genres and Other Late Essays*. Austin: University of Texas Press.

Barnes, Barry, David Bloor, and John Henry. 1996. *Scientific Knowledge. A Sociological Analysis*. Chicago, London: University of Chicago Press.

Blumenthal, Geoffrey. 2013. “On Lavoisier’s Achievement in Chemistry: On Lavoisier’s Achievement in Chemistry.” *Centaurus* 55 (1): 20–47. doi:10.1111/1600-0498.12001.

Butterfield, Herbert. 1949. *The Origins of Modern Science 1300-1800*. New York, London: The Free Press.

Carr, David. 2008. “Narrative Explanation and Its Malcontents.” *History and Theory* 47 (1): 19–30.

Carroll, Noël. 2001. “Interpretation, History and Narrative.” In *Beyond Aesthetics. Philosophical Essays*, 133–56. Cambridge: Cambridge University Press.

Chang, Hasok. 2009. “We Have Never Been Whiggish (About Phlogiston).” *Centaurus* 51 (4): 239–64.

———. 2012. *Is Water H2O? Evidence, Realism and Pluralism*. Boston Studies in the Philosophy of Science. Dordrecht et al.: Springer.

Collins, Harry M. 1994. “A Strong Confirmation of the Experimenters’ Regress.” *Studies in the History and Philosophy of Science* 25 (3): 493–503.

Crombie, Alistair C. 1953. *Robert Grosseteste and the Origins of Experimental Science, 1100-1700*. Oxford: Claredon Press.

Cunningham, Andrew. 1988. “Getting the Game Right: Some Plain Words on the Identity and Invention of Science.” *Studies In History and Philosophy of Science* 19 (3): 365–89.

Cunningham, Andrew, and Perry Williams. 1993. “De-Centring the ‘Big Picture’: The

Origins of Modern Science and the Modern Origins of Science.” *The British Journal for the History of Science* 26 (4): 407–32.

Daston, Lorraine, and Peter Galison. 2007. *Objectivity*. New York: Zone Books.

Feldhay, Rivka. 1994. “Narrative Constraints on Historical Writing. The Case of the Scientific Revolution.” *Science in Context* 7 (1): 7–24.

Franklin, Allan. 1986. *The Neglect of Experiment*. Cambridge: Cambridge University Press.

———. 1994. “How to Avoid the Experimenters’ Regress.” *Studies in the History and Philosophy of Science* 25 (3): 463–91.

Frye, Northrop. 1957. *The Anatomy of Criticism. Four Essays*. Princeton: Princeton University Press.

Giere, Ronald N. 2004. “How Models Are Used to Represent Reality.” *Philosophy of Science* 71 (5): 742–52.

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

Holton, Gerald. 1978. “Subelectrons, Presuppositions, and the Millikan-Ehrenhaft Dispute.” *Historical Studies in the Physical Sciences* 9: 161–224.

Jardine, Nick. 2000. “Uses and Abuses of Anachronism in the History of the Sciences.” *History of Science* 38 (3): 251–70.

Kellert, Stephen H., Helen E. Longino, and C. Kenneth Waters. 2006. *Scientific Pluralism*. Minneapolis: University of Minnesota Press.

Kinzel, Katherina. 2012. “Geschichte Ohne Kausalität. Abgrenzungsstrategien Gegen Die Wissenschaftssoziologie in Zeitgenössischen Ansätzen Historischer Epistemologie.” *Berichte Zur Wissenschaftsgeschichte* 35 (2): 147–62.

Koyré, Alexandre. 1957. *From the Closed World to the Infinite Universe*. Baltimore: John Hopkins University Press.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. 3rd ed. Chicago, London: University of Chicago Press.

Kusch, Martin. forthcoming. “Scientific Pluralism and the Chemical Revolution.” *Studies in History and Philosophy of Science.*

Ladyman, James. 2011. “Structural Realism versus Standard Scientific Realism: The Case of Phlogiston and Dephlogisticated Air.” *Synthese* 180 (2): 87–101.

Lindberg, David C. 1990. “Conceptions of the Scientific Revolution from Bacon to Butterfield: A Preliminary Sketch.” In *Reappraisals of the Scientific Revolution*, edited by David C. Lindberg and Robert S. Westman, 1–27. Cambridge: Cambridge University Press.

McEvoy, John. 2010. *The Historiography of the Chemical Revolution: Patterns of Interpretation in the History of Science*. London: Pickering & Chatto.

Musgrave, Alan. 1976. “Why Did Oxygen Supplant Phlogiston? Research Programmes in the Chemical Revolution.” In *Method and Appraisal in the Physical Sciences. The Critical Backrgound to Modern Science, 1800-1905*, edited by Colin Howson, 181–210. Cambridge, London: Cambridge University Press.

Pomata, Gianna. 1998. “Close-Ups and Long Shots. Combining Particular and General in Writing the Histories of Women and Men.” In *Geschlechtergeschichte Und Allgemeine Geschichte. Herausforderungen Und Perspektiven*, edited by Hans Medick and Anna-Charlott Trepp. Göttingen: Wallstein.

Porter, Roy. 1986. “The Scientific Revolution: A Spoke in the Wheel?” In *Revolution in History*, edited by Roy Porter and Mikuláš Teich, 290–316. Cambridge: Cambridge University Press.

Roth, Paul A. 1988. “Narrative Explanations. The Case of History.” *History and Theory* 27 (1): 1–13.

Sapp, Jan. 1990. “The Nine Lives of Gregor Mendel.” In *Experimental Inquiries: Historical, Philosophical and Social Studies of Experimentation in Science*, edited by H. E. Le Grand, 137–66. Dordrecht: Kluwer.

Secord, James A. 1993. “Introduction to ‘The Big Picture.’” *The British Journal for the History of Science* 26 (4): 387–89.

Shapin, Steven. 1998. *The Scientific Revolution*. Chicago, London: University of Chicago Press.

Tosh, Nick. 2003. “Anachronism and Retrospective Explanation. In Defence of a Present-Centred History of Science.” *Studies in History and Philosophy of Science* 34 (3): 647–59.

Van Fraassen, Bas C. 2008. *Scientific Representation. Paradoxes of Perspective*. Oxford: Claredon Press.

White, Hayden. 1973. *Metahistory. The Historical Imagination in Nineteenth-Century Europe*. Baltimore, London: John Hopkins University Press.

———. 1978. “The Historical Text as a Literary Artifact.” In *The Writing of History. Literary Form and Historical Understanding*, edited by Robert H. Canary and Henry Kozicki, 41–62. Madison, London: University of Wisconsin Press.

———. 1980. “The Value of Narrativity in the Representation of Reality.” In *The Content of the Form. Narrative Discourse and Historical Representation*, 1–25. Baltimore: John Hopkins University Press.

Yates, Frances A. 1964. *Giordano Bruno and the Hermetic Tradition*. Chicago, London: University of Chicago Press.

1. [↑](#footnote-ref-1)
2. Comparable issues arising within the natural sciences have been recently addressed in debates on scientific pluralism (Chang 2012; Kellert, Longino, and Waters 2006). Although I acknowledge that there may exist important parallels between pluralism in science and pluralism in historiography, in this paper I focus on the latter only. Note also that my aim in this paper is to provide a descriptive account of pluralism in historiography. I do not seek to answer the normative question whether pluralism is epistemically desirable or not. [↑](#footnote-ref-2)
3. I use the term “theory” broadly here, such that it includes all sorts of high-level conceptions. These conceptions need not be coherent systematic accounts; indeed, at times it may even be only single assumptions rather than elaborate theoretical constructions that are at issue. Also, I include methodological and basic philosophical commitments about the character of science and the nature of historical change among the sorts of theoretical assumptions that are relevant when talking about theory-ladenness in the historiography of science. [↑](#footnote-ref-3)
4. There is some disagreement as to whether the narrative structure of a historical account carries information, conveys knowledge and is properly representational, or whether narrative is a superfluous, merely „rhetorical“ aspect of historical discourse (Different accounts of the role of narrative in historical representation have been developed by Carr 2008; Carroll 2001; White 1980). I cannot go into these debates here. However, I believe that my analysis in the next section will show that narratives do convey information about the past. In this sense, they should be seen as epistemic rather than „merely rhetorical“. [↑](#footnote-ref-4)
5. Many aspects of White’s narratological account are deeply problematic. On the one hand, his structuralist taxonomy of different modes of historical writing is static, artificial and irritatingly ahistorical. On the other hand, from insights into the constructive dimension of historiography, White draws radical conclusions about its subjective and fictional character. I share neither White’s structuralist inclinations, nor his radical subjectivism, and wish to take from his reflections only the central theses that historical accounts have a narrative structure and that there can exist plural narrative emplotments of the same historical events. [↑](#footnote-ref-5)
6. In the philosophy of science, epistemic criteria such as simplicity, variance of the evidence, surprising predictions, fruitfulness and explanatory power are often thought to help scientists reach a verdict in situations of theory-choice. Since most of these criteria cannot be applied to historiography without difficulties, I develop my account of historiographical evaluation without substantially drawing on discussions of theory-choice in the philosophy of science. [↑](#footnote-ref-6)