Historical Reconstruction:
Gaining Epistemic Access to the Deep Past

Patrick Forber[§] and Eric Griffith[‡]

We discuss the scientific task of historical reconstruction and the problem of epistemic access. We argue that strong epistemic support for historical claims consists in the consilience of multiple independent lines of evidence, and analyze the impact hypothesis for the End-Cretaceous mass extinction to illustrate the accrual of epistemic support. Although there are elements of the impact hypothesis that enjoy strong epistemic support, the general conditions for this are strict, and help to clarify the difficulties associated with reconstructing the deep past.

1. Introduction

Of the many tasks undertaken in science, one is striking both in its scope and the epistemic difficulties it faces: the reconstruction of the deep past. Such reconstruction provides the resources to successfully explain puzzling extant traces, from fossils to radiation signatures, often in the absence of extensive and repeatable observations—the hallmark of good epistemic support. Yet good explanations do not come for free. Evidence can fail, in practice or in principle, to support one hypothesis over another (underdetermination). And when hypotheses do confront conflicting evidence, identifying the piece of theory to abandon can be notoriously difficult (testing holism). Good science, in any discipline, must overcome these challenges.

Our epistemic access to past events is limited, often severely so. Despite the problem of access, some claims about prehistory enjoy strong epistemic support. Here we will analyze the problem of epistemic access for historical reconstruction (§ 2). We will then review a familiar case of successful historical inquiry, the impact hypothesis for the End-Cretaceous mass extinction, to clarify how epistemic support accrues to historical hypotheses (§§ 3–5). We argue that the convergence of independent evidential inferences, a kind of consilience (Whewell 1858), provides the primary source of support for such historical reconstructions. We will argue that the nature of epistemic support is the same or similar across the sciences. The specific strategies for the accrual of epistemic support differ, due to differences in the severity of the epistemic problems different lines of inquiry face.

§ Department of Philosophy, Tufts University, Miner Hall, 14 Upper Campus Road, Medford, MA 02155
‡ Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115
✉ E-mail: patrick.forber@tufts.edu

2. The Problem of Access

The task of historical reconstruction involves crafting a causal etiology for a specific event or set of events.[1] Historical reconstructions provide both a *chronology* and a *history*. The chronology identifies the temporal sequence of events, whereas the history identifies the causal links and processes connecting events across time. Often historical events are unique, and the causal reconstruction can vary in scope from vast (the evolution of the vertebrate eye) to minute (the exact cause of a mechanical failure). We usually cannot conduct repeatable experiments, nor can we observe multiple repetitions. Thus, reconstructing histories and chronologies faces, as does all scientific inquiry, a problem of access.

Limits to epistemic access cause underdetermination problems. An underdetermination problem occurs when available data and various features of competing hypotheses fail to adjudicate a theory choice according to a specific set of decision criteria (Dietrich and Skipper 2007). Underdetermination problems are persistent and pervasive for the task of historical reconstruction. Indeed, much effort is devoted to overcoming these problems. Turner (2005, 220–222) gives several examples from paleobiology and geology where we lack the necessary evidence to discriminate between scientifically serious alternative hypotheses. Forber (2009, 257–263) analyzes an empirically estimated adaptive landscape, revealing how changes to demographic or ecological assumptions about the microevolutionary process alter the landscapes, and thus demonstrates how fragile the evidential assessment of existing data can be..

While all science must confront underdetermination, historical reconstruction often lacks important epistemic recourses available to other lines of inquiry. When reconstructing history we lack the ability to intervene experimentally to test hypothesized causal relationships among events in the past. Moreover, our inability to reproduce or observe repetitions of most historical events ensures that historical reconstruction, unlike tasks involving the identification and testing of regularities, is limited by restricted sources of data. While other areas of scientific inquiry (e.g., celestial mechanics or the structure of the earth) have achieved substantial progress in the face of an inability to intervene experimentally, the task of historical reconstruction faces a further epistemic difficulty: the traces of a past event are subject to disturbance by heterogeneous causal processes over long spans of time, biasing or destroying information extractable from residual traces of the past event. Due to these difficulties, historical reconstruction typically proceeds without sources of replicable data that are insulated from information biasing or destruction. Such data allow science to refine and revise theory to a high degree of precision, to "close the loop" between successive approximations and the phenomena (Smith 2008).

Although the discovery of new residual traces, the advent of new technologies, and the refinement of statistical techniques may all bring new forms of evidence to bear on a historical question, these data are still subject—to varying degrees—to the effects of intervening information-destroying processes. Thus, as Turner (2007) notes, the extractable information of all remaining residual traces may not be sufficient to determine many aspects of past events, causing underdetermination problems. Perhaps of more concern, information destruction raises the possibility that the fragmentary nature of the available traces might be systematically misleading.

The extent of the epistemic limitations introduced by information-destroying processes depends on the detailed circumstances of their operation and on the kind of historical information we aim to collect. In evolutionary biology we know that directional natural selection tends to eliminate variation, and so destroys information about what variants were historically available during the evolution of a lineage. It also tends to move populations to (local) optima, and so destroys information about ancestral trait values and genealogical relationships.[2] But on the molecular level natural selection and common ancestry can leave definite signatures across the genome. For instance, while strong directional selection destroys information about variation and the ancestral trait value on the morphological level, these selective sweeps usually leave detectable signatures in the genome. Of course, the information in sequence data can be destroyed too, and there are limits to how much historical information a sequence can contain. In the extreme case, since there are only four DNA bases, two (neutral) pseudogenes will, after sufficient evolutionary time, be on average 25% similar. At this saturation point evolution has obliterated any extractable information about common ancestry or evolutionary process. Sober and Steel (2002, 398) use information theory to calculate the upper

limit on the information about evolutionary history contained in a dataset. Their framework, or another like it, could be applied to molecular sequences to determine just how far back in time we can reconstruct chronology and history.

How frequently these epistemic limitations create actual underdetermination problems, however, is unclear. Jeffares (2010) provides many reasons to resist such skepticism, for our apparent access to the past continues to improve. To take a single telling example, Turner (2007) cites the coloration of dinosaurs as an illustration of an unanswerable question about the past. In contrast, a recent study claims to identify the plumage color patterns for a late Jurassic theropod dinosaur (Li et al. 2010). While the frequency issue is important, the hard question remains: with no way to test any claim against the actual event in real time, how can epistemic support accrue to historical claims such that they become more than mere just-so stories?

Some (Cleland 2001; 2002; Jeffares 2008), following up a powerful idea discussed in geology, argue that a "smoking gun"—an extant trace that definitively supports one hypothesis over its rivals—provides the main source of epistemic support for historical hypotheses. On this view, a smoking gun constitutes a naturally occurring *experimentum crucis*:

> Take multiple observations of evidence: [o_a, o_b, o_c]. Now take two hypotheses, *H1* and *H2*. If *H1* accounts for [$o_a + o_b$] but is incompatible with [o_c], and *H2* accounts for all three observations [$o_a + o_b + o_c$], then [o_c] is the 'smoking gun' that discriminates between two hypotheses about a historical event. This one downstream effect not only supports one hypothesis, it works against the alternative hypothesis (Jeffares 2008, 471).

This focus on individual lines of evidence makes clear the importance of auxiliary hypotheses but downplays the concomitant problem of testing holism. Connecting an extant trace to a hypothesized event deep into prehistory, back (say) 65 million years to the boundary between the Cretaceous and Paleogene periods, requires many auxiliary hypotheses. These auxiliaries include assumptions about the rates of geological and evolutionary processes, the nature of the fossil record, the calibration of instrumentation, and, as Jeffares (2008) makes clear, applicable causal regularities. Often these indispensable auxiliaries enjoy weak to intermediate epistemic support. And this raises the problem of testing holism. Apparent disconfirmations of a historical hypothesis by extant traces can be explained away by apportioning blame to suspect auxiliaries. More problematically, apparent confirmations of historical hypotheses can be undermined by attacking weaker links in the chain of auxiliaries. Given that any line of evidence will include a number of auxiliaries, and that at least some of the auxiliaries will lack strong epistemic support, there will be a strict limit as to how much epistemic support a smoking gun can provide for a reconstruction.

We will take a different approach, arguing that the main source of epistemic support in historical investigations comes from the consilience of multiple independent lines of evidence on the chronology or key quantitative properties integral to causal history. If lines of evidence that have a high degree of independence yield convergent estimates for the chronology or a quantity assumed by a historical reconstruction, then they provide epistemic support that is less sensitive to testing holism. The independence of evidential inferences is measured by assessing the amount of overlap between the sets of assumed auxiliaries required by the different inferences. If inferences are sufficiently independent, attacking one weak auxiliary will not completely undermine the overall support for the historical reconstruction.

Any analysis of the underlying epistemological basis of scientific inquiry must be carefully rooted in the study of actual scientific practice. To this end, we revisit the impact hypothesis for the Cretaceous-Paleogene (K-Pg) boundary mass extinction.[3] Strong epistemic support derived from smoking guns has been invoked to account for the marked scientific consensus regarding this impact event. Our goal is to reassess the overall epistemic support for the impact, and its proposed causal connection to the cascade of extinctions at the K-Pg boundary, in light of the totality of current evidence. As a result, we pass over several early debates that gradually dissolved as the field matured. We ultimately dispute the view that a smoking gun-based account best captures the underlying epistemic support for this historical hypothesis, arguing that closer inspection reveals that the strong support ultimately derives from the consilience of multiple independent evidential inferences.

3. Reconstructing the K-Pg Mass Extinction

Analysis of the fossilized remains of earlier organisms preserved in geological strata constitutes one of our primary means of epistemic access to the biogeography and demography of prehistoric species. That sedimentary rock strata contain such fossilized flora and fauna, and that these fossils succeed each other vertically in a specific, reliable order that can be identified over wide horizontal distances, was recognized early in the inception of modern geology. Given the so-called law of superposition, the geological principle that sedimentary layers are deposited in a time sequence, with the oldest at the bottom and the youngest on the top, these observations have enabled paleontologists to roughly measure the extinction of species over time. The K-Pg transition marks a period of unusually intense turnover in the nature of preserved fossil types, with an estimated 60-80% of living species and half of the genera disappearing at this boundary (Raup 1988). The historical reconstruction of the K-Pg extinction thus seeks to identify the various contributing causes underlying this anomaly in the pattern of preserved fossilized remains.

The progressive coalescence of robust, independent measures of geological time in the later 20th century established significant new constraints on any acceptable account of the K-Pg transition. Research in this area is currently predicated on the idea that the inferred rapid turnover of species indicates the involvement of processes somehow distinct from those associated with ongoing evolution and extinction. We need to enrich our minimal model of macroevolution to explain this event (Sterelny 2008).[4] As a result, once prominent theories involving the inability of Cretaceous fauna to adapt to slowly emerging environmental pressures, associated with new forms of competition or gradual climate change, have largely given way to hypotheses involving dramatic, geologically-sudden environmental dislocations. In this regard, the attribution of the End-Cretaceous mass extinctions to the effects of the impact of a large earth-crossing asteroid (Alvarez et al. 1980) marked a critical turning point in K-Pg extinction research. Although the possible involvement of extraterrestrial phenomena in the disruption of End-Cretaceous biota had been previously raised by others—notably De Laubenfels (1956) and Russell and Tucker (1971)—the Alvarez hypothesis was unique among such proposals in that it arose from the consideration of distinctive geochemical anomalies detected at the K-Pg boundary layer. Here we will briefly sketch the accrual of epistemic support to the proposal of an End-Cretaceous impact event via the consilience of multiple independent lines of evidence.[5]

The critical initial observation that engendered the impact hypothesis stemmed from a unique collaborative effort between Luis Alvarez, a Nobel prize-winning physicist, and his son Walter, a geologist focused on the K-Pg transition. Working together with geochemists Frank Asaro and Helen Michel, the pair initially set out to measure the iridium concentrations in the thin clay of K-Pg boundary layer sections from the Umbrian Apennines that marked the sudden disappearance of the otherwise abundant fossilized remains of marine plankton. The group hoped to exploit recently observed correlations between sedimentation rate and iridium concentration to determine the length of time represented by this boundary layer. Iridium, a siderophile, is largely absent from the Earth's crust and upper mantle, but is much more abundant in the meteoric dust that constantly accretes from outer space. Thus, assuming a constant accretion rate, the authors hoped to use the concentration of iridium in the sediment to infer the length of time that sediment took to form.

Unexpectedly, the Alvarez team discovered a 30-fold enrichment of iridium (from 0.3 to 9.1 ppb) coincident with the K-Pg boundary layer. Among the 28 elements tested, this effect was specific for iridium, and similar iridium increases were observed at the boundary layers of K-Pg samples from Denmark and New Zealand.[6] The centimeter-thick clay layer would have taken more than a million years to form based on the proposed meteoric dust accretion model, a duration inconsistent with fledgling radiometric measurements of surrounding strata. Interpretations that invoked physical or chemical changes leading to extraction of the iridium resident in seawater also seemed problematic. Such accounts would require seawater concentrations of iridium significantly above those presently observed. Moreover, they suggested that the observed iridium anomaly should be accompanied by a compensating depletion of iridium in the strata immediately above, which was not observed. Thus, the K-Pg iridium anomaly constituted an

intriguing observation, a telltale trace, representing evidence of an unusual event at the time of the extinctions for which none of the then-current hypotheses could easily account.

Despite the scattered distribution of examined K-Pg boundary samples, and an inability to completely rule out the effects of unusual depositional or extractive processes, the Alvarez group adopted a working hypothesis that viewed the iridium anomaly as the result of a sudden influx of extraterrestrial material. They adopted the impact hypothesis rather than a previously advanced scenario involving a nearby supernova (Russell and Tucker 1971) for several reasons. First, iridium enrichment stemming from heavy particle bombardment due to an exploded star should be accompanied by plutonium-244, which was not detected in any of the K-Pg samples. In addition, isotopic analysis of K-Pg clay yielded an ^{191}Ir/^{193}Ir ratio within one percent of that characteristic of solar system material, providing strong independent evidence that the anomalous iridium originated from within the solar system and not from a nearby stellar event. Based on these considerations, the authors concluded that the anomaly arose from the impact of a comet or asteroid, which ejected massive amounts of iridium-rich dust-sized material into the upper atmosphere to be dispersed relatively uniformly around the globe.

Invoking independent data regarding the injection of dust into the stratosphere occasioned by the 1883 Krakatoa eruption, Alvarez and colleagues suggested that crater-derived dust would effectively prevent sunlight from reaching the Earth's surface for a period of several years. This loss of sunlight would suppress photosynthesis, leading to the collapse of terrestrial and marine food chains. Such a scenario, the authors argued, could potentially account for the peculiar taxonomic pattern of K-Pg extinction phenomena. In the open ocean, a temporary absence of sunlight would nearly eliminate photosynthetic algae and similarly disrupt higher levels of the food chain such as foraminifera, ammonites, and marine reptiles. On land, large herbivorous and carnivorous animals that were directly or indirectly dependent on photosynthesizing vegetation would become extinct, whereas smaller vertebrates may have been able to survive by feeding on decaying vegetation or insects, and flora would regenerate from seeds, spores, and existing root systems.

The key geochemical findings of Alvarez's Berkeley team were quickly corroborated by reports from several independent groups (Ganapathy 1980; Kyte 1980; Smit 1980), and provisional adoption of the impact hypothesis inaugurated a new line of research. In the short term, researchers sought to demonstrate that the K-Pg boundary iridium anomaly represented a truly global phenomenon. In the longer term, these findings prompted a reexamination of the K-Pg boundary strata for other residual evidence of a significant impact event—most notably, for the distinctive signature of an impact crater. Specifically, the hypothesis predicted the presence of other geologic phenomena consistent with a large impact event in the K-Pg boundary layer, and that these phenomena would be restricted to this stratigraphic layer.

These efforts quickly bore fruit. Within two years, close to forty K-Pg sites with iridium anomalies were reported worldwide, spanning Europe, Asia, Oceania, Africa, and the Americas. Although more than a decade passed without the identification of an appropriate impact site, the impact hypothesis gained continuing evidential support from its ability to account for a growing number of previously unappreciated and highly unusual geophysical and geochemical features restricted to the K-Pg boundary layer, including:

- Anomalous enrichment of various metals, including osmium, gold, platinum, rhenium, ruthenium, palladium, nickel, and cobalt, in relative proportions typical of their abundance in chondritic meteorites.

- Spherules, unusual rounded glassy rocks that are consistent with the solidified remains of sprayed drops of impact melt.

- Microscopic diamonds consistent with formation through direct shock alteration of solid material at the impact site. Importantly, these minuscule diamonds are inconsistent with a volcanic origin. Given their small size they are unlikely to have survived exposure to the high temperatures of volcanic eruptions. Moreover, they remain almost completely free of nitrogen impurities characteristic of diamonds originating within the Earth's mantle.

- Shocked quartz and stishovite, which only form at pressures greater than 8.5 GPa and have been associated only with suspected meteoroid impact events and nuclear test sites.

- Spinal crystals, nickel-rich metal oxide crystals formed by the melting and rapid solidification of metallic samples in an oxygen-rich environment consistent with the known ablation of meteorites as they enter the Earth's atmosphere.

Thus, by the late 1980s, the idea of a significant K-Pg impact event provided a single account for a diverse array of previously unappreciated geologic and geochemical phenomena, and had encountered no anomalous traces. Despite several false alarms, an appropriately dated impact crater—seemingly the most straightforward implication of the impact hypothesis—remained unidentified for some time. From early on, however, it was recognized that the preservation of this critical piece of evidence would be contingent on its precise location. At the time of the K-Pg transition nearly two-thirds of the surface of the Earth consisted of oceanic crust, of which nearly 25% has since been displaced via subduction, thereby eliminating historical information concerning potential impact craters in these regions. Finally, Hildebrand and colleagues (1991) announced the discovery of a giant impact crater at the K-Pg boundary near the small Mexican fishing village of Puerto Chicxulub.

Following this announcement, evidence quickly mounted in support of the idea that the Chicxulub Crater marks the site of the hypothesized K-Pg impact event. Extensive gravimetric profiling to detect subtle variations in local gravity indicates a crater diameter of 175 to 180 km, matching the estimated impactor size calculated by other methods. These gravimetric findings have been further refined through extensive magnetic and seismic reflection studies (e.g., Hildebrand 1995), and confirmed by direct stratigraphic data from a number of core samples derived via exploratory drilling, all of which display characteristic mineralogical traces of a substantial impact event. In addition, Swisher and colleagues (1992) dated the Chicxulub Crater to 64.98 ± 0.05 mya as compared to a measured date of 65.01 ± 0.08 mya for K-Pg boundary spherules using $^{40}Ar/^{39}Ar$ dating methods. Similarly, melt rock from the Chicxulub core samples and K-Pg impact spherules were found to be isotopically indistinguishable by multi-elemental isotopic analysis, strong support that they have a common origin (Blum 1993).

Given these findings, the scientific community now overwhelmingly endorses the idea of a late Cretaceous impact at Chicxulub. Objectors largely confine their criticisms to issues regarding the fine-grained chronology of the extinctions relative to the Chicxulub impact (e.g., Keller 2004), or seek to challenge the ability of the impact hypothesis to fully account for the selective survival of various species into the Paleogene. In what manner then have researchers overcome the characteristic problem of underdetermination faced by historical reconstructions, and in what respects, if at all, do these various claims differ in their epistemic support? Clearly, the explanatory potential and the promise of more stringently testing the impact hypothesis through further empirical study contributed to the decision to pursue it as a working hypothesis. Here, however, we aim to assess the nature of the current overall epistemic support for the hypothesis. Thus, rather than offering a historical analysis of its acceptance, we focus on the total set of currently available evidence.

4. Overcoming the Problem of Access

In considering the impact hypothesis as an example of successful historical reconstruction, Cleland (2002) has rightly emphasized the crucial importance of the discovery of the anomalous iridium and shocked quartz grains in the widespread acceptance of the account. Given the few known causal mechanisms capable of producing these "smoking gun" phenomena, the ability of the impact hypothesis to unify them under a single consistent causal story serves to discriminate this account from a number of rival hypotheses. Indeed, in speaking of the iridium anomaly, Alvarez and colleagues cite unification as an important virtue of their hypothesis:

> None of the current hypotheses adequately accounts for this evidence, but we have developed a hypothesis that appears to offer a satisfactory explanation for nearly all of the available paleontological and physical evidence (1980, 1095).

Yet, many of the specific claims of the impact hypothesis are supported by a shared, more complex form of evidential reasoning based on the consilience of independent inferences regarding the specific nature and chronology of hypothesized past events.

For instance, consider that while trumpeting the unifying potential for the proposed impact scenario, Alvarez and colleagues also devote a considerable portion of their initial report to the estimation of the size of the hypothesized impactor. One approach based on the size of the observed K-Pg iridium anomaly and the measured abundance of iridium in chondritic meteorites yielded estimated impactor diameters between 6.6 and 14 km. A separate method based on the assumption that the observed 1 cm K-Pg boundary layer is composed of ejected impact material that fell from the stratosphere estimated an impactor diameter of 7.5 km. The authors deem this numerical convergence as significant evidence in favor of the impact scenario, highlighting the fact that several independent estimates of the diameter gave values that lie within the range of 10 ± 4 km.

Although the authors never articulate explicitly the underlying reason for their preoccupation with the impactor's size, their one brief allusion is revealing: "If we are correct in our hypothesis that the C-T [sic] extinctions were due to the impact of an earth-crossing asteroid, there are four independent ways to calculate the size of the object" (Alvarez et al. 1980, 1105). The reasoning here is based on the fact that the impact scenario implies a single defined value for the size of the impacting body. If the iridium anomaly and K-Pg boundary clay do represent residual traces of a common impact event, then to the extent that their properties can be used in conjunction with theory-derived causal relations to infer values for the size of the hypothesized impactor, these values should agree. To be more precise, to the extent that independent inferences from extant data converge onto a single relatively precise range for a hypothesized quantity, they provide strong evidence that there is some definite quantity being measured; the probability that the inferred quantity is indicative of some actual past event or process increases as the number of independent inferential paths taken to the inferred value rises.

Although this pattern of reasoning might be viewed as a subspecies of common cause inference—using various evidential traces to infer a specific property (as opposed to the mere existence) of a common cause— the evidential support provided by this set of inferences possesses features not typically associated with mere unification of the available traces by a common cause. In this form of reasoning the power of a theory to synthesize multiple lines of evidential support is *not* an extra-empirical virtue but a hallmark of the overall degree of support provided by disparate sets of evidence.[7] Moreover, such inferences to properties of the hypothesized cause reinforce each other by drawing upon distinct auxiliary assumptions. The use of several different chains of auxiliaries in the inferences increases the degree of independence, and thus decreases the chance that they are systematically misleading in a way that would yield the same approximate value for the inferred properties of the causal structure associated with the past event. Treating the evidential support resulting from this form of reasoning as mere unification does not explain the role these type of arguments play in science or how they address the problem of testing holism.

These ideas trace back to Whewell (1858, 87–88): "The Consilience of Inductions takes place when an Induction, obtained from one *class of facts*, coincides with an Induction, obtained from a *different class*." For Whewell, consilience—the "jumping together" of inductions based on different sets of facts—provides a crucial source of confirmation for a scientific hypothesis (Snyder 2005). While Whewell predicates his discussion of consilience on an idiosyncratic ontology of classes of facts, we find the general insight illuminating for analyzing the accrual of epistemic support for historical reconstructions. We will take consilience to be the convergence of lines of evidence for specific claims or for values of specific quantities that are parts of a single historical reconstruction.[8] Consilience, a composite accrual of different lines of confirmation-theoretic support, thus constitutes a, if not the, primary measure of support for historical reconstructions of the deep past. While the exact procedure for evaluating consilience is complex and depends on local features of specific evidential inferences, the measure should have the following features.

First, the degree of consilience depends on the extent to which the various inferential paths provide converging constraints on aspects of the hypothesized event history or chronology. Inferences that constrain aspects of the chronology or history in ways that are highly sensitive to the nature of the extant traces provide better support, and help minimize the risk that the apparent convergence arises from the use of mistaken auxiliary assumptions. For each distinct inferential path to the aspect of the reconstruction, we can assess the sensitivity of the inference by asking: if the measured property of the relevant extant trace were altered, how would the inferred value of the hypothesized historical quantity change? The sensitivity of the inference will depend upon both the relevant causal regularities and the permissible range of values for the measured property specified by background theory, and thus may differ for each distinct inferential path to the value of the hypothesized historical quantity. In cases where the inferred quantity is capable of taking a wide range of values and the inference is highly sensitive to the relevant properties of the extant trace (e.g., impactor size or date of impact), even coarse-grained consilience on ranges of the inferred values provides a significant source of epistemic support for a hypothesized reconstruction.

Second, the measure of consilience should also track the degree of independence between the inferential pathways from extant traces to the hypothesized historical reconstruction, where the degree of independence is determined by the amount of overlap between the sets of auxiliary hypotheses invoked by the different paths. Obviously, inferences from extant traces to events in the deep past proceed via various auxiliary hypotheses, including causal regularities and commitments about background conditions (e.g., the typical iridium mass fractional abundance in chondritic meteorites). In addition, such inferences involve a further assumption that the casual system is (approximately) closed. That is, relevant properties of the extant traces have not been significantly altered by the action of extraneous intervening causal processes. For an isolated inferential path, each of these auxiliaries bears the full inferential load—the inferential chain is only as strong as its weakest link. Consilience of multiple independent inferences distributes the epistemic load across auxiliaries employed in multiple inferential paths. As the number of paths to the inferred aspect of the reconstruction increases, the inferential load on auxiliaries that appear in one path lessens insofar as different paths that rely on different auxiliaries converge on the same inferred value. Thus, the support accruing from consilience rises with increasing independence of the inferential paths. The independence between two inferential paths increases as the number of shared auxiliary hypotheses decreases.

The complete causal chain linking present-day traces to the deep past is potentially enormous, and many of these factors are likely to remain unknown. One might therefore question how we can possibly generate a realistic appraisal of the degree of independence. The details matter here. The relevant background theory provides inferential strategies to cope with incomplete information about the causal chain by identifying the possible confounding causal processes. By way of illustration, consider the K-Pg boundary global iridium anomaly as the product of consilient inferences. The impact hypothesis views the iridium increase detected in K-Pg boundary depositions from widely scattered geographic sites as indicative of a common origin: the geologically acute influx of extraterrestrial material in the form of a chondritic meteor. Given that iridium-rich dust from the impact would be expected to persist in the stratosphere for several years and thus distribute relatively uniformly worldwide, each iridium measurement provides an independent way to estimate the quantity of iridium influx, assuming the measured iridium level has not been significantly affected by extraneous processes.

Clearly, a host of intervening processes could have acted to alter extant iridium levels. For a given iridium measurement, a large number of potential intervening mechanisms tend to involve geographically localized phenomena. Thus, with respect to these possibilities the assumption that extraneous processes have not significantly affected the measured iridium level in a given sample is independent of similar assumptions for samples from different locations. While there is some geographic variation among iridium measurements, there is a limited consilience from a wide range of locations that effectively discriminates a feature of the hypothesized historical event (common origin due to impact) from a range of rival possibilities. By contrast, for alternative planetary-scale phenomena (e.g., the rapid extraction of the iridium resident in terrestrial seawater or heavy particle bombardment from a nearby supernova) that might have acted to alter extant iridium levels in K-Pg boundary strata, the same closed system assumption is not independent of similar assumptions for samples from different locations. Different sorts of evidence are necessary to discriminate between rival planetary-scale hypotheses. In the case of the K-Pg iridium anomaly, however, this difficulty is

overcome since evidence of such alternative mechanisms would be expected in currently extant traces and none has been observed (see §3).

Applying these considerations to the four methods discussed by Alvarez and colleagues (1980, 1105–1106) to infer the size of the impacting body, one can see that the stronger inferential approaches are those based on iridium and clay deposition at the boundary layer. Both of these approaches show a high degree of sensitivity. By contrast, the remaining two inferential paths, based on meteor cratering data and light attenuation necessary for an "impact winter," are more conjectural and do not support precise calculations regarding the impactor size. The "impact winter" estimates, for example, provide only a rough lower bound for the value and are tied to a specific scenario in which light attenuation is responsible for the resulting ecological effects. The two sensitive approaches involve a number of independent auxiliary hypotheses. However, they do share common auxiliary assumptions about the fraction of impactor material that persists in the stratosphere and the absence of extraneous effects on the composition of the planetary K-Pg clay boundary layer.[9] None of these methods are particularly strong in isolation, but the consilience on impactor size constrains possible impact scenarios in a way that played a key role in the provisional pursuit of the impact hypothesis.

Subsequent discovery of the Chicxulub crater provided another independent inferential path that supported relatively precise calculations regarding the size of the hypothesized impactor, effectively eliminating any residual debate concerning the origin of the K-Pg iridium anomaly. The inference from the extant traces of the impact crater to the impactor size is derived from experiments and computer simulation studies showing that an impact crater on Earth should be approximately twenty times larger than the impactor's diameter.[10] Moreover, the inference of impactor size from the geophysical features of the Chicxulub site proceeds without reference to the global boundary layer properties, nor does it invoke the behavior of atmospheric impact debris as in the other inferential methods.

By the time of the identification of the impact site at Chicxulub, the notion of a K-Pg impact event had already begun to gain widespread acceptance. This form of inferential consilience, however, is by no means restricted to the estimation of the impactor's size. It was also employed to constrain significantly the date of the impact event. The various geological traces cited as evidence for the impact hypothesis permitted inferences regarding the date of the presumed impact event and those dates all converged to a narrow range. Importantly, these inferred dates derive not just from traditional lithologic stratigraphy, but also involve magnetostratigraphic and radiometric dating methods. Thus, these various methods constitute multiple inferential paths, each with some degree of independence, to date past occurrences.

Analogous approaches were also employed for a number of other quantities associated with the putative K-Pg impact event, including the chemical composition of the impactor, the location of the impact, and the approximate angle of impact. Because the same extant traces can be used to infer the value of several such properties, these inferential ties allow for the progressive integration of disparate forms of evidence into a single unified historical reconstruction in a way that minimizes undue reliance on strong parsimony assumptions. For example, prior to the discovery of the putative impact site at Chicxulub, none of the available data were capable of discriminating between a single impact and multiple smaller, geologically simultaneous impact events. However, the discovery of the putative Chicxulub Crater permitted relatively independent inferences regarding the size of the putative impactor and the date of the impact event that agreed with similar inferences derived from the observable properties of the iridium anomaly. This agreement supported the notion that the putative impact event that produced the iridium anomaly and the impact event that produced the Chicxulub crater were in fact the very same event. As additional research increases the inferential consilience of the available evidence, we garner more and more support for the hypothesis that a single impact event characterized by these various properties actually did take place.

While the resulting consilience of evidential inferences from different traces represents a significant constraint on acceptable historical accounts, any single evidential inference deploys a number of auxiliary hypotheses that lack unequivocal support. This connects to a problem with the smoking gun concept. If we conceive of a smoking gun as a trace that grounds an evidential inference capable of unequivocally supporting one hypothesis over current rivals, we must recognize that there is an inherent limitation on how much epistemic support can accrue from a single trace. Adopting a position, as Cleland does, wherein multiple traces are taken collectively as constituting a smoking gun represents an insufficient remedy.[11] The

primary limitation of such an approach is that it leaves the connections of underlying inferential relations between traces unanalyzed, thereby failing to explicate the nature of the *overall* epistemic support for a historical reconstruction. While it may be the case that "it was widely conceded that the anomalous iridium and shocked quartz provided a 'smoking gun'"(Cleland 2002, 483), this characterization does little to illuminate how the complex inferential connections between these two sets of traces provide epistemic support for the impact hypothesis; the consilience between sets of traces is also epistemologically salient. Moreover, the smoking guns themselves are the product of consilient inductions. Even the multiple anomalous iridium traces only serve as a smoking gun because of the consilience, on both the date and relative magnitude difference of the iridium spikes, among samples taken from around the globe. Thus, the structure of actual epistemic support for the hypothesized K-Pg impact events consists in a complex nested series of overlapping consilient inferences. To the extent that the smoking gun account ignores the nature of these underlying inferential relations between evidential traces, it ultimately understates the evidence in favor of a significant End-Cretaceous impact event.

As many have recognized, claims regarding the exact nature of the impacting body are separate from the central claim concerning the causal connection between the impact and the mass extinction. The peripheral claims about the impact are elaborated and defended in order to support the hypothesis that a significant End-Cretaceous impact event was an actual and major contributing cause of the K-Pg mass extinction. We turn now to the epistemic support for this causal connection. The residual controversy on this point illustrates the epistemic limitations of historical reconstruction.

5. Refining Reconstructions, Uncovering Limitations

The claim that the Chicxulub impact was a significant cause of the K-Pg mass extinction is the same sort of historical claim as those involving the origin of the K-Pg iridium anomaly. Consilience on the chronology and inferred historical properties of the extinction phenomena from independent inferential paths potentially provides the epistemic support for the hypothesized causal relationship between the impact and the extinctions. Alterations in the nature of the putative impact event would imply changes in the ecological pattern or timing of the resulting extinctions, and those would have corresponding implications for the extant fossil record, thereby allowing for two independent inferential paths to the hypothesized extinction phenomena. The first path relies on an analysis of the abundance, biogeography, and morphology of fossilized taxa to assess the ecological changes coincident with the K-Pg transition. The second path utilizes reconstructed parameters of the hypothesized impact event in conjunction with ecological regularities to independently infer the nature of these same late Cretaceous ecological disruptions.

Given the proposed causal relationship, one implication of the impact hypothesis is that impact-induced extinctions should coincide with (or perhaps slightly postdate) the timing of the K-Pg impact event. Some extinctions fit this predicted pattern. The iridium anomaly has been found to coincide stratigraphically with the disappearance of several Cretaceous pollen species (Orth 1981). Similarly, the fossilized remains of normally abundant single-celled foraminifera drop to almost undetectable levels within millimeters of the impact iridium in K-Pg marine strata (Smit 1982). The extinctions of many terrestrial plant species also appear to fit the impact scenario (Nichols and Johnson 2008). In this respect there is good evidence that the impact caused some extinctions.

The impact hypothesis, however, is typically interpreted to extend beyond this narrow claim to include the claim that all (or very nearly all) of the terminal Cretaceous extinctions can be attributed to the consequences of the impact event (Schulte et al. 2010a). Indeed, the impact hypothesis initially attracted so much attention and has been pursued with such vigor because it offered a single mechanism by which to account for the anomalously high number of K-Pg boundary extinctions inferred from the observed fossil record. Yet, *prima facie*, the fossil evidence does not support such a broad claim, showing an uneven disappearance of most taxa from the fossil record *prior* to the K-Pg impact. This incongruity between the impact chronology and the fossil record is problematic because there are alternative causes, such as severe Deccan trap volcanism at the K-Pg boundary, which may have contributed to the mass extinction (Duncan and Pyle 1988; Keller et al. 2008). Controversy over whether the mass extinction has a single cause (impact)

or multiple causes (impact, volcanism, etc.) continues (Archibald et al. 2010; Courtillot and Fluteau 2010; Keller et al. 2010; Schulte et al. 2010b).

Owing to the pervasive effects of information-destroying processes, neither the first nor the last organism in a given taxon will be recorded as a fossil. This raises the question of whether the recorded ranges of fossils can be used to discriminate between gradual and simultaneous extinction hypotheses at a sufficiently fine-grained level of analysis. To address this question Signor and Lipps (1982) built a simple model for fossilization probability based on a Poisson distribution, showing that we should expect significant variation across taxa in the time interval between the last observed fossil and the actual extinction time. The situation here is structurally similar to one discussed by Sober (2008, 318–323) concerning the evidential import of the presence or absence of a fossil intermediate for the question of common ancestry versus separate ancestry. Using Sober's framework, we can formulate the evidentiary situation for observed extinctions in the fossil record as follows.

For a set of two species consider two possible hypotheses: either they share a common catastrophic extinction (*CE*) or they suffered separate gradual extinctions (*SE*). The two hypotheses provide different answers to the question of whether individuals of that species existed at the time of the putative mass extinction event. *CE* entails that they must have existed, while *SE* entails only that they may have (extinction may occur before or after the catastrophe). The situation is complicated, however, by the nature of the fossil record. We can only observe fossils and, due to information destruction, these traces provide only a rough guide for the actual time of extinction. If there is a significant time interval between the last observed fossil (*LOF*) for a species and the mass extinction event (*MEE*), then there is a potential conflict between the separate inferential paths to the extinction chronology. It appears that this extinction fails to cohere with the pattern proposed by *CE*.

Let t_{LOF} be the time (in mya) of the *LOF* and t_{MEE} be the time (in mya) of the *MEE*. For the sake of the argument, assume that after accounting for observational error $t_{LOF} > t_{MEE}$ (i.e., there exists a significant detectable time interval between *LOF* and *MEE*). For a given species, we want to know whether the species existed at some intermediate time t_X, where $t_{LOF} > t_X > t_{MEE}$. For *CE* we know that:

P(species exists at t_X | *CE*) = 1

P(species does not exist at t_X | *CE*) = 0

For *SE* the species may have existed at t_X; let us say this occurs with probability q (directly analogous to Sober's probability q).

$q = P$(species exists at t_X | *SE*)

$1 - q = P$(species does not exist at t_X | *SE*)

Just because a species exists at t_X does not mean that we will be able to search and find a fossil dated t_X. So let us define the probability a (directly analogous to Sober's probability a) as follows:

$a = P$(find fossil dated t_X | *CE* and species exists at t_X) = P(find fossil dated t_X | *SE* and species exists at t_X)

We can now construct a likelihood ratio for the case where we search and fail to find a fossil dated t_X.[12]

P(we do not find a fossil dated t_X | *SE*) / P(we do not find a fossil dated t_X | *CE*) = $(1 - qa)$ / $(1 - a)$

The likelihood ratio has some interesting implications. Given the assumption that $t_{LOF} > t_{MEE}$, searching and failing to observe a fossil from an intermediate time t_X almost always provides evidence for *SE* over *CE*.[13] The strength of evidence (for *SE* over *CE*) depends on the values of q and a. The probability q depends on the exact extinction mechanism presumed by *SE*, but there are some general constraints: q is directly

proportional to $t_X - t_{MEE}$ (i.e., as t_X gets closer to t_{LOF} the probability it exists at that time given SE increases), and we can assume that $q < 1$ because we do not have definitive evidence that the species exists at t_X. The probability a depends on fossilization processes and our observational capabilities, and varies based on a set of species-specific factors, such as physiology, geographical range, and population size. Importantly, for most cases of interest (where $t_{LOF} > t_{MEE}$ and q and a have plausible values for K-Pg extinction phenomena), the choice between SE and CE will be underdetermined.[14]

Assuming differing fossilization probabilities between species, Signor and Lipps demonstrate that in cases of common mass extinctions one should expect significant variation in the time interval between the last observed fossil and the actual extinction time for various species. Similarly, apparently sudden disappearances in the fossil record can be artifacts of geographical range reductions or population bottlenecks. This so-called Signor-Lipps effect provides a means by which the gradual disappearance of taxa from the fossil record prior to the impact date need not be viewed as disconfirming the hypothesis of a simultaneous impact-triggered mass extinction event. Yet this effect also places a significant limitation on the historical reconstruction of this event, entailing that the extant traces of uncommon taxa not represented by a continuous fossil record may ultimately be inadequate to discriminate between gradual and simultaneous extinction hypotheses at a sufficiently fine-grained level of analysis.

Due to their prevalence and reliable fossilization, data on pollen and foraminifera overcome the underdetermination generated by the Signor-Lipps effect—the a values for pollen and foraminiferans are much greater than those for most dinosaurs. Insofar as the inferred extinction dates for these species, derived either from their disappearance from the fossil record or the date of the hypothesized impact event, converge to a single, relatively precise value (Alvarez et al. 1980; Orth et al. 1981; Smit and Hertogen 1980; Smit 1982), this approach provides significant evidence in support of the claim that the impact is an actual contributing cause for at least some End-Cretaceous extinctions. By contrast, the stronger claim that the Chicxulub impact is the major cause of the full scope of K-Pg boundary extinctions receives little direct support from the broader chronology of extinctions inferred from the fossil record. The observed pattern of species loss may be consistent with this hypothesis insofar as fossilized remains of various taxa are not found after the impact event, but underdetermination of the timing of extinction relative to the impact event for many species presents a significant challenge for attempts to obtain high-quality evidence in support of this hypothesis. Ingenious methods to access the potential environmental effects of the impact have uncovered abundant geophysical and geochemical data consistent with the occurrence of tsunamis, wildfires, acid rain, and climate-altering gas emissions expected to result from the hypothesized impact (Kring 2007)—consequences certain to negatively affect a late Cretaceous biota. However, in the absence of a well-supported chronology establishing that many End-Cretaceous extinctions did not take place prior to the K-Pg impact event, these findings do little to support the strong version of the impact scenario.

In general, advocates of the strong version of the impact hypothesis have adopted an alternative approach to support their claims, arguing that the impact scenario is *uniquely* able to account for the breadth and selectivity of the K-Pg extinctions. If we invoke the Signor-Lipps effect to fix the timing of all End-Cretaceous extinctions to the same geologic instant, we are faced with a peculiarly selective mass extinction. The scale of biological turnover is massive, with a number of major animal groups disappearing such as non-avian dinosaurs, marine and flying reptiles, and ammonites (Fastovsky and Sheehan 2005). But crocodilians and avian dinosaurs persisted, and other groups (e.g., benthic foraminifera) showed negligible extinctions (Thomas 2007). In effect, the suggestion that the consequences of the Chicxulub impact can account for the scope and selectivity of these ecological changes amounts to another consilience claim. However, upon close consideration the practical prospects of utilizing the inferred nature of the hypothesized impact event to support detailed reconstructions of the pattern of resulting K-Pg extinctions currently appears dim. Fossil and geochemical evidence do not yet support the necessary fine-grained reconstruction of late Cretaceous ecology. Moreover, we currently lack the theoretical basis necessary to infer the precise nature of the causal relationships between properties of the impact event and specific climatological and ecological consequences (Pierazzo et al. 2003; Kring 2007).

The problem is compounded by the complexity of ecology, and the extreme sensitivity of the effects of the impact on a plethora of environmental and ecological factors is well appreciated:

The environmental consequences of an impact event and any subsequent biological effects rely on several factors, including the ambient environmental conditions and the extant ecosystem structures at the time of impact. Some of the severest environmental perturbations of the Chicxulub impact event would not have been significant in some periods of Earth history. Consequently, the environmental and biological effects of an impact event must be evaluated in the context in which it occurs (Kring 2003, 133).

While advocates of the strong version of the impact hypothesis have appealed to factors such as the potential protective effects of habitat, feeding strategy, body weight, endothermy, and population size to account for the distinctive selectivity of the apparent mass extinction, the insensitivity of the presumed impact scenario to revisions in the pattern of late Cretaceous extinctions (see, e.g., Marshall 1996) betrays the tentative and vague nature of these speculations. Thus, we are still not in a position to utilize geophysical findings to stringently test the role of the impact in the distinctive ecological pattern of K-Pg extinctions. Perhaps this will change with the further resolution of paleoecology and the development of more realistic ecological models of the late Cretaceous, but for now we must acknowledge the severe limitations of this alternative approach.[15]

The limitations uncovered here present a significant difficulty for the strong claim that the impact is responsible for nearly all K-Pg boundary extinctions. In this case the precise, consilient, independent evidential inferences characteristic of the reconstruction of the impact event are absent. Instead, the support for this causal claim rests on a potential circularity: in the absence of evidence to the contrary, the hypothesized impact event is used to fix the timing of extinction for all End-Cretaceous biota, whereas the justification of the impact event as the cause of the extinction phenomena is based on its ability to account for the resulting geologically-instantaneous mass extinction. Given the available evidence for a significant late Cretaceous impact and its apparent effects on various plant and foraminifera species, the strong version of the impact hypothesis is certainly plausible and consistent with the data. However, evidence indicates that other massive geological processes such as catastrophic volcanism and marine regression operated during the late Cretaceous, which may have contributed to the K-Pg biological turnover (MacLeod 2003). Thus, while there is little doubt that an impact played some role in these extinctions, the available data currently fail to discriminate between the strong impact hypothesis and that of multiple smaller, temporally proximate extinction events, or those involving a requisite role for other geological processes. This lingering uncertainty is reflected on the ongoing controversy concerning models in which the Chicxulub impact constitutes the sole cause of the K-Pg mass extinction.

6. Conclusion

We have defended a view of historical reconstruction wherein the consilience of multiple independent lines of evidence on the chronology and key quantities integral to a specific historical reconstruction provides the main source of epistemic support. Unlike the smoking gun account, this approach is specifically sensitive to the problem of testing holism. The consilience of several such inferences involving distinct auxiliary hypotheses onto a relatively precise range of values serves to crosscheck the auxiliaries employed in any one inference, simultaneously testing the legitimacy of all of these background assumptions, and thereby allowing for a stronger conclusion regarding the nature of the inferred past cause of the observed data.[16]

Cleland takes a metaphysical feature of causation to make the smoking gun epistemology viable. Due to the time asymmetry of overdetermination—that events in the causal web tend to have many more downstream effects than upstream contributing causes—historical scientists can:

> proliferate alternative explanations for the traces they observe and then search for a smoking gun to discriminate among them. The overdetermination of the past by the localized present provides the rationale for this work, *ensuring that the probability of finding such traces is fairly high* (Cleland 2002, 494–495, our emphasis).

In effect, Cleland claims that metaphysics makes the probability of successful scientific inquiry into events from the deep past "fairly high" (where success is understood as finding a smoking gun). As stated, this is problematic. Although the proliferation of downstream effects helps and is perhaps a necessary condition for successful reconstruction, the problem is that this metaphysical overdetermination is compatible with epistemic underdetermination (Turner 2005, 215–216). There are many factors that affect, and usually confound, the overall probability of success for scientific inquiry into the past. Processes can destroy information, removing the downstream traces of earlier causes. As the case of fossils clearly shows, the preservation of traces is contingent and often biased. Also, the ability of scientists to find smoking gun traces depends on the instrumentation and background theories at hand, as well as on the specificity of the hypotheses under consideration. Finally, the sheer difficulty of reconstructing the deep past, as evidenced by the history of science, supports the claim that the probability of success is rather low. Causes have many downstream effects, but the probability of successful inquiry also depends on the preservation and detection of those downstream effects.

By this criticism we do not mean to discount the disproportionate role of particular evidential traces in the formulation of historical reconstructions. Because ordinary phenomena can result from the action of any number of distinct causal processes, some anomalous finding indicative of a particular process—a telltale trace—is typically required to gain some inferential purchase on past events. Specifically, a telltale trace helps constrain the possible events or processes capable of producing it, ideally to a small number. Given current background theories, shocked quartz grains, and to a lesser extent the iridium anomaly, constitute such telltale traces. But the strong epistemic support for a historical reconstruction involves much more than telltale traces; it involves integrating aspects of the telltale traces within a chronology and a network of closely interrelated theoretical quantities relevant to the history. This is how the consilience of inferential relations enables such telltale traces to provide epistemic support for the larger reconstructed history of a late Cretaceous Chicxulub impact.

Put more generally, whether some trace counts as a smoking gun depends primarily on historical and sociological context rather than epistemic factors. In particular, a trace gains the status of a smoking gun when the state of research at a time is such that one evidential inference convinces a number of scientists to accept the hypothesis. However, behind the smoking gun is a variety of research that provides multiple consilient lines of evidence that accrue sufficient epistemic support such that one additional line of evidence pushes many scientists over their individual thresholds for acceptance. Thus, although perhaps salient to a historical account of theory acceptance within a particular field, the smoking gun account obscures the nature of the underlying inferential relations between evidential traces that provide epistemic support for a historical reconstruction.

The consilience of multiple inferences to a single parameter is a commonplace strategy used to support theoretical claims in areas outside of historical reconstruction. For example, Salmon (1984, 213–227) argues that the convergence of independent inferences on Avogadro's number was the source of epistemic support that led to the widespread acceptance of the atomic hypothesis in physics. Morgan (1934) made an argument for the reality of genes based on a similar sort of evidential convergence that fits the kind of consilience we defend here. Janssen (2002) provides several additional cases from the history of science where a general consilience style of reasoning is used to provide epistemic support for scientific claims.

While consilience provides a good source of epistemic support for claims pursued in any line of scientific inquiry, conditions for accruing support for historical reconstructions through consilience are strict. Not only does the approach depend on the availability of extant data that can serve as inputs for necessary auxiliaries to invoke causal generalizations that are projectable back in time, but the data must also support multiple independent inferential paths to various hypothesized quantities associated with these past events. Furthermore, because this method is grounded on the presumption that inferences based on mistaken assumptions will not yield the same value for the parameter of interest, the epistemic power of this approach is only realized in cases where the different inferential lines have a sufficient degree of independence, and yield estimated values and chronologies associated with the historical reconstruction that are highly sensitive to the precise nature of the extant traces. Nor can the consilience of multiple lines of inference be taken as sufficient support to definitively establish a given historical reconstruction. This strategy can be progressively

undermined by the need to appeal to coincidence—invoking a precise set of historical conditions that are unsupported by independent lines of evidence in order to account for all available evidentiary traces. The strong version of the impact hypothesis that must appeal to a precise, if ill-defined, set of environmental conditions to account for the hypothesized pattern of impact-related ecological disruptions that purportedly caused almost all the K-Pg extinctions exemplifies this appeal to coincidence. Likewise, the presence of anomalous traces undercuts epistemic support garnered by the consilience strategy in the absence of a mitigating account of the phenomena. For example, the disappearance of various taxa from the fossil record prior to the hypothesized impact event is a *prima facie* anomalous trace, but the Signor-Lipps effect largely mitigates this anomaly by accounting for differences between fossil dates and actual extinction dates. Finally, in isolation, the consilience approach does little to address the possibility that information-destroying processes have rendered the available fragmentary traces systematically misleading. The threat of such a possibility must be assessed on a case-by-case basis.

Detailed reconstruction of the deep past is difficult, as evidenced by the history of science. Even in cases where the relevant causal processes fall within well-defined theoretically mature domains, historical reconstruction often proves extremely challenging. A line of historical inquiry overcomes the persistent problem of access to achieve substantial epistemic authority not only by identifying one or more anomalous telltale traces, but also by isolating a substantial body of high-quality data that supports multiple independent inferences about the properties of past events in the hypothesized causal network. Given these challenges, it is perhaps not unexpected that the pace of historical reconstruction in some contexts lags behind scientific inquiry into phenomena that occur regularly or repeatedly.

Literature cited

Alvarez, L.W., Alvarez, W., Asaro, F., and H.V. Michel. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction. Science 208: 1095–1108.

Alvarez, W. 1997. *T. rex* and the Crater of Doom. Princeton: Princeton University Press.

Alvarez, W. 2003. Comparing the evidence relevant to impact and flood basalt at times of major mass extinctions. Astrobiology 3: 153–161.

Archibald, J.D., Clemens, W.A., Padian, K., et al. 2010. Cretaceous extinctions: multiple causes. Science 328: 973.

Beatty, J. and E.C. Desjardins. 2009. Natural selection and history. Biology & Philosophy 24: 231–246.

Blum, J.D., Chamberlain, C.P., Hingston, M.P., Koeberl, C., Marin, L.E., Schuraytz, B.C., and V.L. Sharpton. 1993. Isotopic comparison of K/T boundary impact glass with melt rock from the Chicxulub and Manson impact structures. Nature 364: 325–327.

Cleland, C.E. 2001. Historical science, experimental science, and the scientific method. Geology 29: 987–990.

Cleland, C.E. 2002. Methodological and epistemic differences between historical science and experimental science. Philosophy of Science 69: 474–496.

Courtillot, V. and F. Fluteau. 2010. Cretaceous extinctions: the volcanic hypothesis. Science 328: 973–974.

De Laubenfels, M.W. 1956. Dinosaur extinction: one more hypothesis. Journal of Paleontology 30: 207–218.

Dietrich, M. and R.A. Skipper. 2007. Manipulating underdetermination in scientific controversy: The case of the molecular clock. Perspectives on Science 15: 295–326.

Duncan, R.A. and D.G. Pyle. 1988. Rapid eruption of the Deccan flood basalts at the Cretaceous/Tertiary boundary. Nature 333: 841–843.

Fastovsky, P. and P.M. Sheehan. 2005. The extinction of dinosaurs in North America. GSA Today 15: 4–10.

Forber, P. 2009. Spandrels and a pervasive problem of evidence. Biology & Philosophy 24: 247–266.

Frankel, C. 1999. The End of the Dinosaurs: Chicxulub Crater and Mass Extinctions. New York: Cambridge University Press.

Ganapathy, R. 1980. A major meteorite impact on the Earth 65 million years ago: evidence from the Cretaceous-Tertiary boundary clay. Science 209: 921–923.

Harper, W. 1989. Consilience and natural kind reasoning. In: An Intimate Relation. Ed. by J.R. Brown and J. Mittelstrass. Dordrecht: Kluwer.

Henderson, L., Goodman, N.D., Tenenbaum, J.B., and J.F. Woodward. 2010. The structure and dynamics of scientific theories: A hierarchical Bayesian perspective. Philosophy of Science 77: 172–200.

Hildebrand, A.R., Penfield, G.T., Kring, D.A., Pilkington, M., Camargo, Z.A., Jacobsen, S.B., and W.V. Boynton. 1991. Chicxulub crater: a possible Cretaceous/Tertiary boundary impact crater on the Yucatan peninsula, Mexico. Geology 19: 867–871.

Hildebrand, A.R., Pilkington, M., Connors, M., Ortiz-Aleman, C. and R.E. Chavez. 1995. Size and structure of the Chicxulub crater revealed by horizontal gravity gradients and cenotes. Nature 376: 415–417.

Janssen, M. 2002. COI stories: Explanation and evidence in the history of science. Perspectives on Science 10: 457–522.

Jeffares, B. 2008. Testing times: regularities in the historical sciences. Studies in the History and Philosophy of Biological and Biomedical Sciences 39: 469–475.

Jeffares, B. 2010. Guessing the future of the past. Biology & Philosophy 25: 125–142.

Keller, G., Adatte, T., Gardin, S., Bartolini, A., and S. Bajpai. 2008. Main Deccan volcanism phase ends near the K-T boundary: evidence from the Krishna-Godavari basin, SE India. Earth and Planetary Science Letters 268: 293–311.

Keller, G., Adatte, T., Pardo, A., Bajpai, S., Khosla, A., and B. Samant. 2010. Cretaceous extinctions: evidence overlooked. Science 328: 974–975.

Keller, G., Adatte, T., Stinnesbeck, W., Rebolledo-Vieyra, M., Fucugauchi, J.U., Kramar, U. and D. Stüben 2004. Chicxulub impact predates the K-T boundary mass extinction. Proceedings of the National Academy of Sciences USA 101: 3753–3758.

Kring, D.A. 2003. Environmental consequences of impact cratering events as a function of ambient conditions on Earth. Astrobiology 3: 133–152.

Kring, D.A. 2007. The Chicxulub impact event and its environmental consequences at the Cretaceous-Tertiary boundary. Palaeogeography, Palaeoclimatology, Palaeoecology 255: 4–21.

Kyte, F.T., Zhou, Z., and J.T. Wasson. 1980. Siderophile-enriched sediments from the Cretaceous-Tertiary boundary. Nature 288: 651–656.

Li, Q., Gao, K.-Q., Vinther, J., Shawkey, M.D., Clarke, J.A., D'Alba, L., Meng, Q., Briggs, D.E.G., and R.O. Prum. 2010. Plumage color patterns of an extinct dinosaur. Science 327: 1369–1372.

Lyell, C. 1868. Principles of Geology. London: John Murray.

MacLeod, N. 2003. The causes of Phanerozoic extinctions. In: Evolution on Planet Earth. Ed. by L. Rothschild and A. Lister. London: Academic Press.

Marshall, C.R. and P.D. Ward. 1996. Sudden and gradual molluscan extinctions in the latest Cretaceous of western European Tethys. Science 274: 1360–1363.

Morgan, T.H. 1934. The relation of genetics to physiology and medicine. In: Nobel Lectures, Physiology or Medicine 1922-1941. Amsterdam: Elsevier Publishing Company (1965).

Myrvold, W. 2003. A Bayesian account of the virtue of unification. Philosophy of Science 70: 399–423.

Nichols, D.J. and K.R. Johnson. 2008. Plants and the K-T boundary. Cambridge: Cambridge University Press.

Orth, C.J., Gilmore, J.S., Knight, J.D., Pillmore, C.L., Tschudy, R.H., and J.E. Fassett. 1981. An iridium abundance anomaly at the palynological Cretaceous-Tertiary boundary in northern New Mexico. Science 214: 1341–1343.

Pierazzo, E., Hahmann, A.N., and L.C. Sloan. 2003. Chicxulub and climate: radiative perturbations of impact-produced S-beary gases. Astrobiology 3: 99–118.

Powell, J.L. 1998. Night Comes to the Cretaceous: Comets, Craters, Controversy, and the Last Days of the Dinosaurs. New York: W.H. Freeman and Company.

Raup, D.M. 1988. Extinction in the geologic past. In: Origins and Extinctions. Ed. by D.E. Osterbrock and P.H. Raven. New Haven: Yale University Press.

Russell, D. and W. Tucker. 1971. Supernovae and the extinction of the dinosaurs. Nature 229: 553–554.

Salmon, W. 1984. Scientific Explanation and the Causal Structure of the World. Princeton: Princeton University Press.

Schulte, P., Alegret, L., Arenillas, I., et al. 2010a. The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. Science 327: 1214–1218.

Schulte, P., Alegret, L., Arenillas, I., et al. 2010b. Response. Science 328: 975–976.

Signor, P.W. and J.H. Lipps. 1982. Sampling bias, gradual extinction patterns, and catastrophes in the fossil record. In: Geological Implications of Impacts of Large Asteroids and Comets on the Earth. Ed. by L.T. Silver and P.H. Schultz. Geological Society of America Special Publication.

Smeenk, C. 2003. Approaching the Absolute Zero of Time: Theory Development in Early Universe Cosmology. PhD thesis, Dept. of History and Philosophy of Science, University of Pittsburgh.

Smit, J. 1982. Extinction and evolution of planktonic foraminifera after a major impact at the Cretaceous/Tertiary boundary. In: Geological Implications of Impacts of Large Asteroids and Comets on the Earth. Ed. by L.T. Silver and P.H. Schultz. Geological Society of America Special Publication.

Smit, J. and J. Hertogen. 1980. An extraterrestrial event at the Cretaceous-Tertiary boundary. Nature 285: 198–200.

Smith, G.E. 2008. Closing the loop: testing gravity, then and now. In: The Isaac Newton Lectures at the Suppes Center. Stanford University.

Snyder, L.J. 2005. Consilience, confirmation, and realism. In: Scientific Evidence: Philosophical Theories and Applications. Ed. by P. Achinstein. Baltimore: John Hopkins University Press.

Sober, E. 1988. Reconstructing the Past: Parsimony, Evolution, and Inference. Cambridge, MA: MIT Press.

Sober, E. 2008. Evidence and Evolution: The Logic Behind the Science. New York: Cambridge University Press.

Sober, E. and M. Steel. 2002. Testing the hypothesis of common ancestry. Journal of Theoretical Biology 218: 395–408.

Stegenga, J. 2009. Robustness, discordance, and relevance. Philosophy of Science 76: 650–661.

Sterelny, K. 2008. Macroevolution, minimalism, and the radiation of animals. In: The Cambridge Companion to the Philosophy of Biology. Ed. by D. Hull and M. Ruse. Cambridge: Cambridge University Press.

Swisher, C.C., Grajales-Nishimura, J.M., Montanari, A., Margolis, S.V., Claeys, P., Alvarez, W., Renne, P., Cedillo-Pardoa, E., Maurrasse, F.J-M.R., Curtis, G.H., Smit, J., and M.O. McWilliams. 1992. Coeval $^{40}Ar/^{39}Ar$ ages of 65.0 million years ago from Chicxulub crater melt rock and Cretaceous-Tertiary boundary tektites. Science 257: 954–958.

Thomas, E. 2007. Cenozoic mass extinctions in the deep sea: What perturbs the largest habitat on earth? Geological Society of America Special Papers 424: 1–23.

Turner, D. 2005. Local underdetermination in historical science. Philosophy of Science 72: 209–230.

Turner, D. 2007. Making Prehistory: Historical Science and the Scientific Realism Debate. Cambridge: Cambridge University Press.

Turner, D. 2009. Beyond detective work: Empirical testing in paleontology. In: The Paleobiological Revolution: Essays on the Growth of Modern Paleontology. Ed. by D. Sepkoski and M. Ruse. Chicago: University of Chicago Press.

Weinberg, S. 2008. Cosmology. New York: Oxford University Press.

Whewell, W. 1858. Novum Organon Renovatum. London: John W. Parker.

Notes

1. The term "historical reconstruction" should be understood broadly to encompass causal investigations into both individual spatiotemporally compact past events (e.g., a particular extraterrestrial impact event) and temporally extended sequences of past events (e.g., determining the effects of an extraterrestrial impact on the late Cretaceous flora and fauna). This differs from another typical task in science: the identification of regularities (laws of nature or otherwise). The latter treats specific events as instances of general patterns, rather than as targets of investigation. Some, notably Cleland (2001, 2002), distinguish between two different prototypical approaches to science: historical and experimental. However, it would be mistaken to view these distinctive scientific tasks as mapping directly onto traditional scientific disciplines, for most have both experimental and historical aspects. As

Jeffares (2008) persuasively argues, paradigmatic historical sciences—such as paleobiology, archaeology, and geology—make extensive use of regularities, and often construct models with the aim of identifying new regularities. Turner (2009) makes the case that part of paleobiology involves testing distribution hypotheses about large-scale patterns rather than causal reconstructions. And paradigmatic experimental sciences often engage in historical reconstruction. Early universe cosmology, for example, focuses on the reconstruction of the first few minutes of the universe, and such reconstruction cannot be disentangled from the theoretical innovations and identified regularities in current particle physics (Smeenk 2003; Weinberg 2008).

2. Sober (1988, 3–4) made this point early on. Systems with multiple local equilibria preserve more information about initial conditions than those with a single global equilibrium. Beatty and Desjardins (2009) continue this sort of analysis and reveal further complications for reconstructing evolutionary history by investigating the sensitivity of natural selection to other evolutionary factors.

3. Formerly known as the Cretaceous-Tertiary (K-T) extinction, the Tertiary period has since been divided into the Paleogene and Neogene periods.

4. In the interest of brevity we background several factors here. Attributing disappearance from the fossil record to species extinction represents a significant inference beyond the available data, displayed in cases such as coelacanths. Moreover, even granting that disappearance from the fossil record might be reasonably equated with extinction, the disappearance of many taxa within a specific level of the sedimentary record need not necessarily imply rapid extinction in a manner that could not be accounted for solely by basal evolution. Lyell (1868), for example, viewed such discontinuities in the fossil record as mere gaps in the sedimentary record, where long periods of time had not been recorded, or were erased by erosion, and thus gave the mistaken impression of abrupt radical change. Thus, from a uniformitarian perspective, to the extent to which the spike in apparent extinction intensity at the K-Pg transition demands explanation, the account to be provided concerns alterations in the sedimentation rate, not alterations in the extinction rate. Only with the relatively recent establishment of robust independent measures of geological time were these underdetermination problems overcome.

5. For more detailed historical accounts of this research and the surrounding controversy, see Alvarez (1997), Powell (1998), and Frankel (1999).

6. Iridium concentrations from the three samples showed substantial variation (from 9.1 ± 0.5 ppb in the Italian samples to 41.6 ± 1.8 ppb in the Danish samples). However, the measured values all represented ≥ 20-fold increases over background levels observed in surrounding strata.

7. However, see Myrvold (2003) for a modified Bayesian model that renders unification as an empirical virtue. Also, see Stegenga (2009) for a more skeptical view on syntheses of multiple sources of evidence.

8. Turner (2007) discusses the role of consilience in historical sciences, but his treatment differs from ours. For Turner, consilience is a kind of explanatory unification, and while it may carry some weight in theory choices (Turner suggests that it can be used to "break evidential ties between rival hypotheses"), we should remain skeptical about consilience because there will often be "equally consilient" rivals to choose between (2007, 203). We want to avoid equating consilience with simple unification, for we see consilience as a thoroughly empirical virtue—it is the synthesis of multiple lines of evidential support, and therefore an empirical measure. In addition, treating consilience as unification does not explain the role consilience-type arguments play in addressing the problem of testing holism. Along these lines, although not explicitly discussing the independence of contributing inferences, Janssen (2002) provides an account of consilience as the combination of multiple common cause inferences, or "common origin inferences." He argues that these "meta-common origin inferences" have played a crucial role in the history of science, discussing Copernicus, Kepler, and Poincaré on the rotation of the Earth as well as styles of argument found in Newton, Whewell, and Darwin. See also Harper (1989) for an analysis of consilience in Newton's argument for universal gravitation. Furthermore, consilience considerations extend to inductive methods beyond common cause inferences. Consider the importance of estimating quantities relevant to historical reconstruction with greater and greater precision (e.g., the K-Pg impactor size, specific chronological dates using radiometric methods, dates of evolutionary divergence using molecular methods, etc.).

9. The two approaches share an estimate for the fraction of impact material that initially persisted in the stratosphere. Termed the "Krakatoa fraction" after the eruption that provided an estimate of the parameter, the value 0.22 was used despite the different character of the two explosions because it was the only relevant number then available. However, because of the way this term enters into the two inferences, the inferred impactor sizes obtained with the two methods converge regardless of the specific value employed. Thus, with respect to the consilience obtained using these two approaches, the shared assumption only consists in the more modest claim that the behavior of

iridium-rich ejected impact material in the atmosphere approximates that of other impact ejecta. Once we consider the convergence of inferred impactor sizes using these approaches with that obtained using Chicxulub cratering data, however, the specific value employed from the Krakatoa fraction again becomes relevant.

10. The inferred impactor diameter varies linearly with the size of the measured impact crater, yielding an estimated impactor diameter of 9-10 km.

11. Cleland explicitly endorses the idea of multiple traces jointly constituting a smoking gun. As Cleland (2002, 480–481) defines it: "A smoking gun is a trace(s) that unambiguously discriminates one hypothesis from among a set of currently available hypotheses as providing 'the best explanation' of the traces thus far observed." For the impact hypothesis, she views the combination of the iridium anomaly and shocked quartz as a smoking gun.

12. Following standard confirmation theories, we will assume that the observations confirm the hypothesis that confers the higher likelihood on such observations over alternative hypotheses that confer lower likelihoods.

13. The only exception to this generalization involves extreme cases where $q' = 1$ or $a' = 0$. This entails a likelihood ratio of 1 and so the data have no evidential import in these cases.

14. For illustration, suppose *SE* entails $q = 0.1$ and that our fossil finding methods entail $a = 0.1$; then the likelihood ratio is a mere 1.1. Given all the factors that affect q and a, this is weak evidence indeed. Notice that a varies across species due to factors such as population size, geographical range, or fossilization potential.

15. The sensitivity of the impact-induced ecological effects to a host of environmental and ecological factors also undermines another strategy sometimes invoked to support the strong version of the impact hypothesis: that the K-Pg boundary exemplifies a general causal regularity linking significant impact events to terrestrial mass extinction phenomena. In this regard, while scattered traces do connect impact events to periods of rapid species turnover, any correlation between these phenomena remains weak (Alvarez 2003). Moreover, even advocates of such a correlation readily grant that evidence for a potential causal relationship is strongest in the case of the K-Pg boundary extinctions.

16. Rich, complex Bayesian models of epistemology, such as hierarchical models (Henderson et al. 2010) or modified unification models (Myrvold 2003), provide a point of departure for constructing a formal model of consilience, which should help further clarify the role of consilience in science.