Local Underdetermination in Historical Science*

Derek Turner†‡

David Lewis (1979) defends the thesis of the asymmetry of overdetermination: later affairs are seldom overdetermined by earlier affairs, but earlier affairs are usually overdetermined by later affairs. Recently, Carol Cleland (2002) has argued that since the distinctive methodologies of historical science and experimental science exploit different aspects of this asymmetry, the methodology of historical science is just as good, epistemically speaking, as that of experimental science. This paper shows, first, that Cleland's epistemological conclusion does not follow from the thesis of the asymmetry of overdetermination, because overdetermination (in Lewis's sense) is compatible with epistemic *under*determination. The paper also shows, contra Cleland, that there is at least one interesting sense in which historical science is epistemically inferior to experimental science, after all, because local underdetermination problems are more widespread in historical than in experimental science.

1. Introduction. David Lewis (1979) describes several different kinds of asymmetries between the past and the present, on the one hand, and the present and the future, on the other. Most importantly, he argues that counterfactual dependence is asymmetrical, meaning that while the future is counterfactually dependent on the present, the past is counterfactually independent of the present. Lewis argues that the time asymmetry of counterfactual dependence explains a number of other interesting time asymmetries, including the time asymmetry of causation (causes always precede their effects) and the time asymmetry of openness (the future looks open to us, but the past looks fixed). At the end of the article, Lewis goes on to suggest that *overdetermination* is asymmetrical as well. Carol Cleland (2002) has recently invoked Lewis's thesis of the time asymmetry

*Received January 2004; revised June 2004.

†To contact the author, please write to: Department of Philosophy, Connecticut College, 270 Mohegan Avenue, New London, CT 06320; e-mail: derek.turner@conncoll.edu.

‡I thank several anonymous reviewers for this journal whose constructive comments helped me to improve the paper. I also wish to thank Kate Kovenock, Michael Lynch, and Brian Ribeiro for letting me try some of these arguments out on them.

Philosophy of Science, 72 (January 2005) pp. 209–230. 0031-8248/2005/7201-0012$10.00
Copyright 2005 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/426851 Published online by Cambridge University Press

of overdetermination in order to answer the charge that prototypical historical science is epistemically inferior to classical experimental science. Cleland argues that the asymmetry of overdetermination is a fact about our universe that underwrites the distinctive methodologies of historical and experimental science, guaranteeing that the one methodology is, epistemically speaking, just as good as the other. In this paper, I argue that Lewis's notion of the asymmetry of overdetermination cannot do the work that Cleland wants it to do. I also give a reason for thinking that historical science is, in at least one interesting sense, epistemically inferior to experimental science.

2. Lewis on the Asymmetry of Overdetermination. Lewis defines the 'determinant' of any fact about the world as "a minimal set of conditions, jointly sufficient, given the laws of nature, for the fact in question (Members of such a set may be causes of the fact, or traces of it, or neither)" (1979, 474). A fact or affair is overdetermined just in case it has more than one determinant at a given time. Overdetermination, Lewis suggests, is a matter of degree. A fact may have two or three determinants, or many more.

There are some familiar examples of earlier affairs overdetermining later affairs. For instance, when a convict is shot by firing squad, the death is overdetermined. Only one shot would have been sufficient to kill him. For another example, suppose that Jones shakes hands with three different people, all of whom have the flu. Jones' getting sick is overdetermined by these three handshakes. Although such cases show that earlier affairs sometimes overdetermine later affairs, Lewis thinks that cases like this are uncommon. Moreover, in most of these cases, the number of determinants is quite small. On the other hand, Lewis argues that overdetermination of earlier affairs by later affairs is both more common and more extreme: "*We may reasonably expect overdetermination toward the past on an altogether different scale from the occasional case of mild overdetermination toward the future*" (Lewis 1979, 474; my emphasis). Call this highlighted claim the "thesis of the time asymmetry of overdetermination." Notice that this is strictly a metaphysical thesis. There are a couple of different questions we might ask about this thesis. First, is it true? Second, does it have any interesting epistemological consequences?

Is overdetermination really asymmetrical, as Lewis suggests? Lewis's thesis is initially plausible. Suppose, for example, that two different people throw baseballs at the same window at the same time. In that case, the shattering of the window is overdetermined. Cleland points out that the breaking of the window is overdetermined by numerous subcollections of shards of glass lying on the kitchen floor (2002, 487). That overdetermination of earlier facts by later traces occurs whenever a window breaks.

https://doi.org/10.1086/426851 Published online by Cambridge University Press

Beyond pointing to examples like this, it is not clear to me how one would go about defending (or, for that matter, criticizing) such a thesis. This is especially true, because Lewis's thesis admits of exceptions: he only claims that earlier affairs *seldom* overdetermine later affairs, and that later affairs *usually* overdetermine earlier affairs.

For purposes of this paper, I will assume that Lewis's thesis of the asymmetry of overdetermination is true. I shall argue, however, that Cleland is wrong to think that this metaphysical thesis has any interesting epistemological consequences. More specifically, she is wrong to suppose that the asymmetry of overdetermination "underwrites" the distinctive methods of prototypical historical science and classical experimental science, in the sense of guaranteeing that neither methodology is epistemically better than the other. Lewis, I think, hints that this thesis has no epistemological consequences when he says that "Most of these traces are so minute or so dispersed or so complicated that no human detective could ever read them" (1979, 474).

3. Cleland's Argument. According to Cleland, the methods of prototypical historical science differ from those of classical experimental science. Historical scientists proceed in roughly the following way:

1. Observe and describe puzzling traces of long-past events.
2. Postulate a common cause of those traces. The common cause is usually some token event or process that occurred long ago.
3. Test this hypothesis about the distant past against rival hypotheses by searching for a "smoking gun," or a present trace that, together with the other traces observed so far, is better explained by one of the rival hypotheses than by the other. (2002, 481)

A smoking gun does not necessarily provide support for a hypothesis considered independently of rival potential explanations. Rather, as explanatory hypotheses proliferate, scientists search for smoking guns that will discriminate among them. Cleland gives a number of convincing examples of smoking guns in historical science. For instance, the presence of iridium and shocked quartz at the Cretaceous-Tertiary boundary, not to mention the Chicxulub crater in Central America, are smoking guns for the Alvarez hypothesis that an asteroid impact triggered the extinction of the dinosaurs (Alvarez et al 1980). None of the other potential explanations of the Cretaceous-Tertiary mass extinction imply the existence of a crater.

Cleland shows that historical scientists exploit the asymmetry of overdetermination in the following way: The thesis of the asymmetry of overdetermination implies that most events in the past will have a large number of determinants at the present time, where each determinant is

https://doi.org/10.1086/426851 Published online by Cambridge University Press

a set of conditions (or traces) that, together with the laws of nature, are jointly sufficient for the earlier event. This explains how the distinctive methodology of historical science can deliver scientific knowledge of the past.

Next, Cleland contrasts prototypical historical science with classical experimental science. She emphasizes that since these are ideal types, a particular piece of scientific work may be partly historical, partly experimental. Historians often reason experimentally, and experimentalists sometimes reason historically. According to her, practitioners of classical experimental science (an ideal type) proceed in the following way:

1. Begin by forming a hypothesis about a regularity among event types.
2. Predict what will happen if the hypothesis is true, and if a given test condition is realized.
3. Run a series of experiments in which conditions are manipulated so as to rule out false positives and false negatives.

For example, suppose that an ecologist wants to test a hypothesis about the effects of deer browsing on local vegetation. She makes a prediction about what sorts of plants would grow in a given spot, were they not browsed by deer, and she tests this hypothesis by fencing off a small plot of forest and waiting to see what happens. The ecologist then repeats the experiment while varying certain conditions, such as the amount of sunlight available to the plants or the acidity of the soil, by fencing off different plots in different places—for example, one on top of a dry ridgeline and another in a shady ravine.

Cleland argues that the experimental method is an attempt to cope with or even circumvent the time asymmetry of overdetermination. In order to make this point about experimental science, she relies on the following example: A short circuit is not sufficient for the occurrence of a destructive fire. It is only a partial cause. In order for the fire to occur, there must be flammable materials nearby, the sprinkler system must malfunction, and so on. The burning down of the house is therefore causally *under*-determined by the short circuit.

Suppose that after the ecologist fences off a certain patch of woods to prevent deer browsing, saplings of a given tree species begin to flourish in the protected area. The ecologist still needs to consider other possible causal influences. What if some other animal, aside from the deer, has been destroying the saplings? And what if the fence also succeeds in keeping out that other animal? In that case, the experimental results might yield a false positive. Thus, "there is a need to ferret out and control for additional factors that are relevant to the total causal situation," and that is just what experimentalists do when they manipulate test conditions (2002, 494).

https://doi.org/10.1086/426851 Published online by Cambridge University Press

Cleland's argument, then, can be summarized as follows:

P1. Later affairs usually overdetermine earlier affairs, but earlier affairs usually underdetermine later affairs.
P2. Historical scientists exploit one half of this asymmetry: their methods for testing hypotheses about past event tokens are appropriate because later affairs usually overdetermine earlier affairs.
P3. The experimental method is a strategy for coping with the other half of this asymmetry: since earlier affairs and events (such as the short circuit) usually underdetermine later affairs and events (such as the burning down of the house), anyone who wishes to test hypotheses about regularities among event types must run a series of trials in which different test conditions are manipulated, with the aim of ruling out false positives and false negatives.
C. Therefore, prototypical historical science and classical experimental science are equally good, epistemically speaking.

This is an ingenious argument, and Cleland does an excellent job making the case for P2 and P3. I think she is probably right that historical and experimental science exploit different aspects of the time asymmetry of overdetermination, and by pointing this out, she has contributed a great deal to our understanding of the relationship between the two methodologies. My goal in this paper is to explain why the epistemological conclusion does not follow from the premises.

At certain points in her paper, Cleland shifts from talking about overdetermination as defined by Lewis to talking about epistemic overdetermination. She says, for example, that "the asymmetry of (epistemic) overdetermination is ultimately founded on a time asymmetry of nature" (2002, 489). She also says that "the overdetermination of causes by their effects is (strictly speaking) only epistemic" (2002, 488). As we have seen, Lewis's thesis about the asymmetry of overdetermination is a metaphysical one. For Lewis, the determinant of any affair can be either a set of earlier causes, or a set of later traces, and the determination relation is a relation among affairs (or facts), not a relation between hypothesis and evidence. The suggestion I wish to make is that Cleland is misled into thinking that the conclusion follows from the premises stated above because she fails to distinguish clearly between causal/metaphysical overdetermination of the sort that Lewis is talking about and epistemic overdetermination.

What is epistemic overdetermination? One initially plausible suggestion is that a hypothesis or theory H is epistemically overdetermined just in case there at least two distinct arguments, or lines of evidence, each of which alone is sufficient to justify believing H. For example, someone (not me) might think that belief in the existence of God is epistemically overdetermined by the various arguments for God's existence, because

https://doi.org/10.1086/426851 Published online by Cambridge University Press

each of those arguments, by itself, would be sufficient to justify belief in God. I am not sure if this is what Cleland means by epistemic overdetermination. At any rate, I will argue that the asymmetry of overdetermination does not imply an asymmetry of epistemic overdetermination, in this sense of epistemic overdetermination. On the contrary, metaphysical overdetermination, in Lewis's sense, is compatible with epistemic *under*determination.

At one point in her paper, Cleland says that the asymmetry of (causal? epistemic?) overdetermination could be probabilistic:

> Although Lewis characterizes the asymmetry of overdetermination in terms of sufficiency, it could turn out to be a probabilistic affair, with the ostensibly overdetermining subcollections of traces lending strong but, nevertheless, inconclusive support for the occurrence of their cause. Like the determinism in Lewis's original version, the probabilistic support offered by collections of traces for hypotheses would be an objective feature of the world. (2002, 490)

Elsewhere she refers to this probabilistic phenomenon as "the asymmetry of (quasi) overdetermination" (2002, 491). I have four distinct worries about this passage. First, it is hard to tell whether this probabilistic overdetermination is an epistemic or a causal notion. The claim that it is "an objective feature of the world" suggests a causal/metaphysical notion, as in Lewis's original version. However, the reference to "probabilistic support" suggests an epistemic notion. Second, how is quasi-overdetermination different from quasi-underdetermination? Why use the word 'overdetermination' at all where we are not talking about sufficiency, as in Lewis's original version? Third, what reason is there to think that this probabilistic quasi-overdetermination is asymmetrical? Perhaps earlier affairs quasi-overdetermine later affairs to the same extent that later affairs quasi-overdetermine earlier ones. It is possible that quasi-overdetermination is asymmetrical, but further argument is needed to support this new thesis. Finally, there are still plenty of nontrivial cases of local epistemic *under*determination in which later traces do not even lend probabilistic support to hypotheses about earlier events.

In order to see why causal/metaphysical overdetermination does not imply epistemic overdetermination, and why it is compatible with epistemic underdetermination, we need only look at a case in which there is both metaphysical overdetermination and epistemic underdetermination. I will use one of Cleland's own examples.

4. Why Causal/Metaphysical Overdetermination Does Not Rule Out Epistemic Underdetermination. Cleland uses the example of a baseball shattering a window in order to illustrate Lewis's thesis of the asymmetry of

https://doi.org/10.1086/426851 Published online by Cambridge University Press

overdetermination. The baseball hitting the window does not overdetermine the later traces (i.e., the shards of glass landing on the kitchen floor), but there are many subcollections of traces that overdetermine the baseball's hitting the glass, in Lewis's sense of overdetermination.

Now suppose we develop the thought experiment a bit further. The owners of the house sweep up the shards, toss the baseball in the closet, and eventually repair the window. A few weeks later, the only traces of the event that remain are a few shards of glass underneath the refrigerator. The housecleaning and repair are examples of what Sober (1988, 3) calls *information-destroying processes*. Consider the epistemic situation of the historical investigator who finds the shards of glass under the refrigerator. The investigator may grasp that they are traces of some sort, without having any idea what they are traces of. Are the shards the remains of a broken window, a broken wine glass, or a broken picture frame? Even if the historical investigator recognizes the traces for what they are, rival hypotheses about earlier events and processes will often be underdetermined by the available traces. After studying the shards under the refrigerator, the historical investigator will be completely stymied: The evidence does not permit her to discriminate at all between incompatible rival hypotheses (window vs. wine glass, football vs. baseball, etc.). Moreover, since the investigator knows that people usually clean up the mess when things like windows and wine glasses break, she has good reason to think that she will never find any traces that will enable her to distinguish between the rival hypotheses. In other words, she confronts a local epistemic *under*determination problem.

Or does she? Cleland might point out that the processes of cleanup and repair will leave traces of their own—a receipt for the window filed away somewhere, tiny pieces of glass stuck in the bristles of the broom, and so on. This is true, but unhelpful. Suppose the historical investigator finds some small bit of evidence suggesting that a window was shattered. Did a football or a baseball do the damage? Instead of allowing the researcher to investigate the scene a few weeks after the fact, make the investigator wait for a few decades, until all the traces of the cleanup have been cleaned up, scattered, or destroyed.

What this shows, I think, is that Lewis's thesis of the asymmetry of overdetermination does not rule out epistemic underdetermination. This is precisely the sort of case in which the "traces are so minute or so dispersed or so complicated that no human detective could ever read them" (Lewis 1979, 474). Lewis's thesis of the asymmetry of overdetermination is compatible with the epistemological thesis that local underdetermination problems are widespread in historical science, and that is what I will show in Section 5. Indeed, I will argue that there is reason to think that local epistemic underdetermination is a bigger problem in his-

https://doi.org/10.1086/426851 Published online by Cambridge University Press

torical than in experimental science, and so there is reason to think that Cleland's conclusion (C, above) is false.

Cleland's example of the Chicxulub crater, which is a smoking gun for the hypothesis that an asteroid collided with the earth approximately 65 million years ago, is typical of historical science in one way, but not in another. It is typical of historical science as far as methodology goes, because the scientists in this case sought to test their hypothesis by finding a smoking gun, just as Cleland describes. But it is atypical of historical science as far as epistemology goes. The event in question was of such a magnitude, and happened so recently (65 million years is not so long ago, geologically speaking) that its presently observable traces are a dead giveaway, just as the shards of glass and the baseball on the floor would be a dead giveaway to any investigator who happened on the scene before the homeowners had repaired the window and cleaned up the mess. It would be a mistake to infer from this sort of example that earlier causes are usually, or even very often, epistemically overdetermined by their effects.

One potential objection at this point is that the example of the baseball shattering the window is misleading because it involves human agency. One might reasonably think that in nature, there is no one to "clean up after" geological events, and nothing analogous to the person who repairs the broken window. Why use a hypothetical scenario involving human agency when we are mainly interested in prehistory? I have several responses to this worry: First, the example of the ball shattering the window does show that overdetermination of earlier affairs by later affairs (in Lewis's sense) is compatible with epistemic underdetermination, which is all that I have aimed to show so far. Second, the example is Cleland's own. I hope to have shown that even in the case that she herself uses to illustrate the time asymmetry of overdetermination, earlier events can be (epistemically) underdetermined by their later traces. Third, information-destroying processes in nature erase historical traces just as cleanup and repair erase the traces of the collision of the baseball with the window. Whether or not the traces are destroyed as a result of human agency is inessential to the argument. What matters is that our background theories give us reason to believe that they have been destroyed.

5. Local Underdetermination Problems in Historical Science. Historical scientists frequently find themselves in situations similar to that of the investigator who discovers a few shards of glass under the refrigerator. In order to show this, I will begin by offering an analysis of local epistemic underdetermination problems; then I will describe four cases from historical science that fit the analysis.

Two incompatible theories or hypotheses, H and H^*, are empirically

https://doi.org/10.1086/426851 Published online by Cambridge University Press

equivalent in the *weak* sense if and only if they are both equally well supported by all the available evidence. By contrast, H and H^* are empirically equivalent in the *strong* sense just in case they are (or would be) equally well supported by all the empirical evidence that will ever be available to us.

There are plenty of trivial cases of strong empirical equivalence in historical science. For example, what color were the dinosaurs? On page 138 of David Norman's popular book, *The Prehistoric World of the Dinosaur* (just one example among many), there is a picture of a gray pachycephalosaur with a neon blue patch on the top of its head. At the beginning of the book, Norman writes that it is "difficult—in fact, almost impossible—to know what colors dinosaurs were" (1988, 8). Why only almost? Norman adds that it is possible to make guesses based on analogies with living organisms. For example, most big herbivores—elephants, rhinoceroses, and hippopotamuses—have dull grayish colors. Perhaps the same was true of the big herbivores of the Mesozoic. Nevertheless, anyone can see that the hypothesis that *Pachycephasosaurus* had a neon blue patch on its head is strongly empirically equivalent to the hypothesis that it had a neon green patch. We have good reason to think this because we know that information about coloration is destroyed by the fossilization process. Our background theories of taphonomy tell us that we will never find any historical traces that render either of these hypotheses more probable than the other.

Can we be sure that the rival hypotheses about the color of *Pachycephalosaurus* are strongly empirically equivalent? Suppose that in the future we encounter an extraterrestrial civilization that sent a zoological expedition to earth many millions of years ago to conduct a detailed survey of the earth's flora and fauna, and that these extraterrestrials possess color photographs of pachycephalosaurs. One might reasonably argue that since we cannot rule out this possibility, we cannot be sure that the rival hypotheses about the color of *Pachycephalosaurus* are strongly empirically equivalent. However, we still have no reason at all to think that we ever will come to possess such photographs. Furthermore, our background theories about taphonomy do give us good reason for thinking that all the information about the colors of dinosaurs has been completely destroyed, and therefore considerable justification for thinking that rival hypotheses about the color of *Pachycephalosaurus* are strongly empirically equivalent. In this case, although we cannot be certain that H and H^* are strongly empirically equivalent, because we cannot rule out the possibility of an encounter with alien dinosaurologists, our background theories nevertheless give us good reason for thinking that the rival hypotheses are strongly empirically equivalent.

https://doi.org/10.1086/426851 Published online by Cambridge University Press

A *local underdetermination problem* is any situation in which the following conditions are satisfied:

a. Two incompatible hypotheses, *H* and *H**, are genuine rivals.
b. *H* and *H** are weakly empirically equivalent.
c. As best anyone can tell, *H* and *H** have roughly equal portions of nonempirical theoretical virtue (simplicity, explanatory power, and the like).
d. Background theories give us some reason to think that *H* and *H** are also strongly empirically equivalent.

When these conditions are met, scientists ought to suspend judgment with regard to *H* and *H**. This could mean that they continue to search for a smoking gun that will discriminate between the two, even if there is no good reason to think that such a smoking gun will ever turn up. Or it could mean that they simply move on to more tractable research questions. It is easy to see how philosophers of science could underestimate the pervasiveness of local underdetermination problems in historical science—as I think Cleland does—because scientists themselves tend not to dwell on such problems. For this reason also, examples of local underdetermination problems in historical science are likely to seem a little contrived. No serious scientist would spend time looking for a smoking gun to distinguish between rival hypotheses about the colors of the dinosaurs, because there is good reason to doubt the existence of any such clues. Indeed, historical scientists are trained to identify local underdetermination problems and to move on to more tractable research questions. For this reason, it would be difficult to produce examples of research problems that scientists are currently working on, and that clearly satisfy the above conditions for a local underdetermination problem.

Much of the discussion of underdetermination has focused on what might be called the *global underdetermination problem*. The global problem is generated by the empirical equivalence thesis that for any hypothesis *H*, there is at least one strongly empirically equivalent rival. Some philosophers have even suggested that for any hypothesis *H*, there are indefinitely many strongly empirically equivalent rivals. In fact, it is easy to show that the following is true:

> *Strong Historical Empirical Equivalence Thesis*. For any hypothesis about the past *H*, there are indefinitely many strongly empirically equivalent rivals.

We can generate the rivals algorithmically, in the following way: Consider the hypothesis that God created the universe at some past time *t* (six seconds ago; six minutes ago; six thousand years ago; six trillion years ago, etc.), and that when he did so, he made the universe to appear older/

https://doi.org/10.1086/426851 Published online by Cambridge University Press

younger than it really is. Since there are indefinitely many past times that we can plug in for *t*, we can form indefinitely many creationist hypotheses, each of which will be strongly empirically equivalent to any other historical hypothesis we care to dream up (see also the algorithms proposed by Kukla 1996).

Philosophers of science with naturalist leanings commonly react to this radical skepticism about the past by dismissing it on the grounds that it is not the sort of epistemological problem that arises during the course of actual scientific research. P. Kyle Stanford expresses this feeling well when he says that "underdetermination was supposed to represent a distinct and important problem, arising perspicuously in the context of scientific theorizing about inaccessible domains of nature and troubling *even those who never hoped to defend their scientific beliefs to the truly radical skeptic*" (2001, S3; his emphasis). One popular way of dismissing the global underdetermination problem is to argue that the algorithmically generated hypotheses, such as the hypothesis that God created the world a few minutes ago, are not "genuine rivals" of any scientific hypotheses.

This points to one interesting difference between local and global underdetermination problems. Whereas local underdetermination problems arise during the course of scientific inquiry, global underdetermination problems are imposed upon science by philosophers. Anyone who thinks that the algorithmically generated rivals deserve to be taken seriously will not find the local underdetermination problems to be very interesting or important, simply because global underdetermination is stronger than local. Let us suppose, then, if only for the sake of argument, that naturalistic philosophers are correct to dismiss the hypothesis that God created the world a mere five minutes ago on the grounds that it is not a genuine rival of any scientific hypothesis about the past. This supposition opens the door for a serious consideration of local underdetermination problems. However, it also means that subsequent conclusions will be conditional upon this being the right response to the global underdetermination problem.

Condition *c* in my analysis of local underdetermination problems may need some further clarification. Some philosophers have sought to appeal to the nonempirical theoretical virtues in order to block the inference from empirical equivalence to evidential equivalence. If H and H^* are strongly empirically equivalent, it might still be reasonable to prefer H if we could show that it is simpler, or that it has more explanatory power than H^*. This move raises a number of notorious problems: First, how are we to define 'simplicity' and 'explanatory power' with any precision? Second, why should we suppose that these desirable features are epistemic as opposed to merely pragmatic virtues? What reason is there to think that they are reliable indicators of truth, or approximate truth, or likelihood?

https://doi.org/10.1086/426851 Published online by Cambridge University Press

(For discussion of some of these problems, see Kukla 1994). I make no attempt to address these problems here. It is worth emphasizing, however, that even if these problems could be solved, the appeal to nonempirical theoretical virtue would not necessarily break an evidential tie between strongly empirically equivalent hypotheses, because it is possible neither H nor H^* affords a simpler or better explanation than the other. This point is frequently overlooked, because philosophers tend to focus more on global than on local underdetermination problems.

Now for the cases:

(i) *Caytonia* is an extinct gymnosperm from the Mesozoic era. The name was originally given to fossilized reproductive organs consisting of two rows of ovules attached to a central stalk. Two other kinds of structures have since been found. The first are longish pollen-bearing structures. Palynologists suspect that the pollen-bearing structures belong to *Caytonia* because the pollen found in them matches that found in the fossilized ovules. Both reproductive structures are, in addition, associated with clusters of three to six leaflets attached to the end of a stalk. There is no shortage of fossilized parts, and we know that *Caytonia* plants were fairly widespread in Mesozoic North America. The challenge, then, is to infer the architecture of the whole plant on the basis of these fossilized parts. To this day no one knows whether *Caytonia* plants were trees, vines, shrubs, or herbs, and probably no one ever will (Cleal and Tomas 1999, 95).

(ii) Ever since the Reverend Edward Hitchock began cataloging and describing fossil footprints in the Connecticut River valley in the 1830s, vertebrate paleontologists have had the problem of reconciling two distinct taxonomic systems: the familiar system based on skeletal remains and a system of ichnotaxa based on trace fossils, such as footprints. What exactly is the relationship between dinosaur ichnotaxa, such as *Eubrontes* and *Grallator*, and the more familiar taxa that have been identified on the basis of skeletal remains? One problem is that the parataxonomy based on fossil footprints is coarser-grained than the taxonomy based on skeletal remains. Since background theories of taphonomy tell us that the conditions most conducive to the preservation of skeletons in the fossil record are completely different from the conditions most favorable for the preservation of footprints, nearly every fossil trackway poses an underdetermination problem: how can we tell which sort of animal made this particular set of tracks? Footprint fossilization typically happens under the following conditions: A deep pond recedes during a dry season, leaving fine-grained bottom sediments exposed to the air. An animal that comes to the pond to drink walks across the muddy flat, leaving a set of footprints. In the days and weeks that follow, the exposed sediment containing the trackway is baked in the sun and hardened. Then at some later point

https://doi.org/10.1086/426851 Published online by Cambridge University Press

the rains come, the pond is flooded once again, and a new layer of coarser sediment blankets the old, filling in the tracks. If this new layer of sediment hardens in the right way, the footprints will be preserved in the bedding planes of the resulting sedimentary rock. Whereas footprints need to spend some time baking in the sun in order to be preserved, rapid burial, as in a flash flood, is most conducive to the preservation of teeth and bones (Thulborn 1990). We may know that a given set of tracks was made by a sauropod (see, e.g., Wilson and Carrano 1999), and even, if they are wide-gauge, that they were made by a titanosaur, but several titanosaur genera have been identified based on skeletal remains. Were the tracks made by *Saltasaurus* or by *Titanosaurus*? We will probably never know the answers to such fine-grained questions.

(iii) Jenkins (2000) has criticized the snowball earth hypothesis, advanced by Kirschvink (1992) and Hoffman et al. (1998) in order to explain evidence of low-latitude glaciation during the Neoproterozoic, approximately 800–580 million years ago, by proposing a rival hypothesis that also explains the glacial debris. According to the snowball earth hypothesis, the entire planet was covered by a layer of ice and snow, on several occasions during the Neoproterozoic, for several million years at a time. Jenkins argues that if the earth's obliquity, or the tilt of its axis, had been different during the Neoproterozoic than it is today, then low latitudes might have received less energy from the sun than the higher latitudes. Localized glaciation near the equator is just what one would expect to see if the earth's tilt exceeded 54°. The available evidence does not discriminate between the snowball earth hypothesis of global glaciation and the hypothesis that radical climate changes, including local glaciation at the equator, occurred during the Neoproterozoic as a result of changes in the earth's obliquity. Evans (2000) tried to distinguish between these two rival hypotheses by looking for evidence of glacial deposits in regions that would have been near the poles during the Neoproterozoic, assuming the high-obliquity hypothesis is correct, but he found none.

(iv) Finally, rival adaptationist hypotheses are all too frequently underdetermined by the available evidence. An adaptationist hypothesis is a hypothesis about what a given trait or behavior is (or was) an adaptation *for*. For example, Barlow (2000, chapter 1), citing the work of Janzen and Martin (1982), argues that the fruit of the avocado tree is an evolutionary anachronism. The avocado tree coevolved with the Pleistocene megafauna of Central and South America: gomphotheres, giant ground sloths, and the like. These animals were large enough to swallow avocadoes, carry the pits around in their guts, and later deposit them far from the parent tree. Barlow speculates that the oily green avocado fruit was originally an adaptation for attracting these seed dispersers, all of which have been extinct for approximately eleven thousand years. Since

https://doi.org/10.1086/426851 Published online by Cambridge University Press

that time, humans have been the main dispersers of avocado seeds. Barlow also points out that jaguars are known to eat whole avocadoes in the wild, and agoutis gather and bury them just as squirrels bury acorns, but she says that "the fruit of the avocado was not shaped by millions of years of selection for these underabundant, ill-fitted or fickle dispersal agents" (2000, 11). But how can we be sure? Was the fruit of the avocado tree an adaptation for attracting ground sloths? Or an adaptation for attracting gomphotheres? While it may be reasonable to suppose that the oily flesh of avocadoes was an adaptation for attracting seed dispersers, hypotheses about *which* fauna were the main dispersers—sloths, jaguars, rodents, gomphotheres, some or all of the above?—are underdetermined by the available evidence.

All four of these cases satisfy condition *a*, because in all of them the rival hypotheses are produced by scientists in the course of scientific investigation. They will all count as genuine rivals on any reasonable account of rivalry or theoreticity. Second, *b*, the rival hypotheses in each of the four cases are weakly empirically equivalent. The other two conditions deserve a bit more attention.

First, it is relatively uncontroversial that the first two cases satisfy condition *c*.

H. *Caytonia* was a shrub.

*H**. *Caytonia* was a vine.

If anyone were to define 'simplicity' or 'explanatory power' in such a way as to yield the result that either of these hypotheses is simpler, or affords a more powerful explanation than the other, I think that alone would be good enough reason to reject the proposed definition. Things may not be quite so simple in cases (iii) and (iv). It is conceivable that someone could show that the high-obliquity hypothesis is simpler than the snowball earth hypothesis, or that the snowball earth hypothesis explains more. In the absence of any precise definition of 'simplicity' or 'explanatory power', all we can do is to make impressionistic judgments. Sober's (1988) treatment of the notion of cladistic parsimony shows one way in which philosophers of science can give precise definitions of such notions. According to Sober, when a scientist appeals to simplicity to break a tie between competing hypotheses, that appeal is a surrogate for stating a well-confirmed background theory. For example, cladists' appeals to the notion of parsimony are just disguised appeals to background assumptions about the nature of evolutionary processes (1988, 64–65). Anyone who wishes to challenge my claim that these four cases satisfy condition *c* will need to do something analogous to what Sober has done with the notion of cladistic parsimony, and then show, once simplicity has been clearly de-

https://doi.org/10.1086/426851 Published online by Cambridge University Press

fined, that the snowball earth hypothesis (for example) is simpler than the high-obliquity hypothesis.

Do all four cases satisfy condition *d*? Do relevant background theories give us any reason to suspect that *H* and *H** are strongly empirically equivalent? I will argue in the next section that the answer is yes.

Before going on to develop the main argument of the paper, however, I want to address one potential worry about these four examples: None of them are examples of the underdetermination of scientific *theories*. Stanford (2001, S5–S6) suggests that the only really convincing examples of empirically equivalent theories come from physics. To be sure, the well-known historical theories of biology and geology (such as Darwin's evolutionary theory, plate tectonics, and so on) are not underdetermined at all in the sense that I am using here. However, the fact that the four cases I have described are not examples of theoretical underdetermination does not make them any less interesting. Local underdetermination problems such as (i) through (iv) arise during the course of Kuhnian "normal" historical research. The distinction between larger-scale theoretical underdetermination and smaller-scale underdetermination of hypotheses will not matter for the argument of this paper.

6. How Historical Processes Destroy Information. There is one very general reason for thinking that local underdetermination problems are more pervasive in historical than in experimental science. Background theories of geology, and especially taphonomy, tell us that many historical processes—the fossilization process, the processes of weathering and erosion, continental drift, subduction, glaciation, and so on—are *information-destroying processes*, rather like housecleaning and document shredding. Elliott Sober (1988, 3–5) uses the following example to illustrate this concept of an information-destroying process. Suppose a person releases a ball from the rim of a giant bowl. A later observer happens along and finds the ball resting at the bottom of the bowl. It will be impossible for the observer to infer from which point along the rim the ball was released. No one hypothesis about the point of release is any more probable than another. In this case, rival hypotheses about the point of release are underdetermined by the observable evidence, because all of them are empirically equivalent in the strong sense. The interesting thing about the example, however, is that we have background knowledge (of bowls, gravity, and so forth) that leads us to expect that rival hypotheses will be empirically equivalent in the strong sense. We can even explain how and why the process by which the ball rolls to the bottom of the bowl destroys information about the point from which it was released.

The situation in prototypical historical science is analogous. Kemp (1999) describes various kinds of incompleteness in the fossil record: For

https://doi.org/10.1086/426851 Published online by Cambridge University Press

instance, *biogeographic incompleteness* is a serious problem for paleoecologists. Suppose that a population of terrestrial animals migrates seasonally between dry upland areas and wetter lowland areas that are well drained by rivers. Conditions in dry upland areas are not well suited to fossilization, so it is a good bet that the only members of this species who make it into the fossil record will be the ones that die in lowland areas, along river banks or in floods. This means, however, that the fossil record will give us a distorted picture of the range of these animals. Another upshot of this is that some biological communities will be far more extensively represented in the fossil record than others. The point is simply that we know that the fossilization process destroys information about biological communities in dry upland areas. Another relevant problem discussed by Kemp is that of *stratigraphic incompleteness*, which arises because sediments do not accumulate at a constant rate. The periodic flooding of a major river, such as the Mississippi, affords a good example of this. Since more sediment is deposited during floods than at other times, when scientists look at a layer of sedimentary rock, they are looking at sediments that accumulated in fits and starts. Kemp points out that if there were a period during which no sediment was deposited, that can have a distorting effect on the fossil record. Suppose that some population of organisms living in the neighborhood was evolving at a steady rate during a given stretch of time. Suppose, further, that a large amount of sediment was deposited during the early part of this stretch, and a lot of sediment was deposited during the later part, with a lengthy gap in between, during which time the local rivers, for whatever reason, happened not to flood. This stratigraphic incompleteness will create the illusion of rapid, or even punctuated, evolutionary change in the population. Scientists looking at the record will not be able to discriminate between the hypothesis of gradual evolutionary change during a time in which no sediment accumulated, or the hypothesis of steady sedimentation and rapid evolutionary change.

I conclude that condition d is satisfied in all four of the nontrivial cases described above, which means that they are all bona fide cases of local underdetermination. In all four cases, our background knowledge of the incompleteness of the geological record gives us at least some reason to think that the rival hypotheses are strongly empirically equivalent.

Laudan and Leplin (1991) suggest that there might also be some general reasons for thinking that H and H^*, though weakly empirically equivalent, are not strongly empirically equivalent. First, the empirical consequence class of any hypothesis is determined, in part, by auxiliary assumptions that are subject to revision over time. It is at least possible that paleontologists will revise some of the background assumptions of taphonomy—the very background theories that, for the moment, give us reason to

https://doi.org/10.1086/426851 Published online by Cambridge University Press

think that *H* and *H** are strongly empirically equivalent—and if this were to happen, it could turn out that *H* and *H** are not strongly empirically equivalent at all. This point is well taken, and it is one reason why we should be careful about jumping to the conclusion that any pair of rivals, *H* and *H**, are strongly empirically equivalent. However, there is no reason to think that our background theories of taphonomy are going to be significantly revised anytime soon, and since those background theories do provide some reason for thinking that the rival hypotheses in these four cases are strongly empirically equivalent, it is correct to describe these as cases of local epistemic underdetermination. It is also worth pointing out that Laudan and Leplin's main target is the global underdetermination argument; their observation about the instability of auxiliary assumptions is compatible with the existence of local underdetermination problems, such as I have described.

Laudan and Leplin (1991) also emphasize that the range of the observable is liable to change, which is another reason why we should hesitate to conclude that weakly empirically equivalent rivals are also strongly empirically equivalent. This point, too, is well taken. Suppose that engineers devise a new fossil detection gizmo that enables scientist to study fossils buried in places that are otherwise inaccessible. It is possible that the new gizmo would enable scientists to find a "smoking gun" that would discriminate between the hypothesis that *Caytonia* was a vine and the rival hypothesis that it was a shrub. But this sort of consideration is not terribly helpful. Scientists have found loads of partial *Caytonia* fossils, suggesting that the conditions favorable to fossilization of the leaves and reproductive structures were, for whatever reason, unfavorable to the preservation of the other parts of the plant. Based on this track record, there is some reason to doubt that the smoking gun is even out there for us to find.

In sum, Laudan and Leplin's arguments show that it would be rash to assert that the rival hypotheses in these four cases are strongly empirically equivalent, but that is not what d asserts. Condition d only says that background theories about historical processes lend some support to the claim that *H* and *H** are strongly empirically equivalent.

7. A Fossilized Dinosaur Heart. What about cases in which, contrary to what our background theories may lead us to believe, someone *does* find a smoking gun that discriminates among hypotheses that once looked to be strongly empirically equivalent? Consider, for instance, the question whether or not dinosaurs were endothermic. A few decades ago, it might have been reasonable, given the available background theories, for scientists to conclude that no one will ever find a smoking gun to lend support to one or the other hypothesis. In other words, a few decades ago, hy-

https://doi.org/10.1086/426851 Published online by Cambridge University Press

potheses about dinosaur metabolism may well have constituted a local underdetermination problem, according to the above analysis. However, in recent years, scientists have discovered a number of different historical traces—including an apparent fossilized dinosaur heart that has four chambers and one aorta, just like a bird's—that clearly support the hypothesis that dinosaurs were endothermic (Fisher et al. 2000). This example seems to show that not all local underdetermination problems in historical science are permanent.

Yet there are two reasons for thinking that this serendipitous smoking gun is one of those exceptions that proves the rule. First, some respected scientists have doubted that the object which Fisher et al. studied using CT scans is a fossilized heart at all. Rowe, McBride, and Sereno (2001) point out that the alleged fossilized heart was found inside the chest cavity of a *Thescelosaurus* sekeleton, in the sandstones of the Hell Creek Formation in Montana. They argue, on the basis of taphonomy, that it is highly implausible to suppose that the internal organ of a dinosaur could ever have been preserved in such sedimentary environments. They suggest that Fisher et al. were in fact looking at an ironstone concretion and not at a fossil at all, for "ironstone concretions are notorious for producing suggesting and misleading shapes," and they have often been found in conjunction with dinosaur bones in the American west (2001, 783a). Regardless of the eventual outcome of this debate, it is instructive to see that in this case, specialists are arguing *from* background theories about information-destroying processes *to* the conclusion that what seems like a stunning example of a smoking gun may not be a smoking gun at all. The background theories of taphonomy are that powerful.

Second, suppose that the object scanned by Fisher et al. really is a fossilized dinosaur heart, and that it is a serendipitous smoking gun. If so, that only gives rise to new research questions, and—arguably—new local underdetermination problems. For example, Fisher et al. point out that *Thescelosaurus* is an ornithischian ("bird-hipped") dinosaur, whereas modern birds, with their four-chambered hearts, are thought to be more closely related to the saurischian ("lizard-hipped") dinosaurs. Did the four-chambered heart evolve independently in several dinosaur lineages, or did it evolve early on in dinosaur history, perhaps even before the saurischians and ornithischians diverged? This, as Fisher et al. point out, "remains an open question." Since there is no reason to expect that we will find any more fossilized dinosaur hearts, the answers to such questions about the evolution of dinosaur hearts will probably remain locally underdetermined. Thus, even if it is a genuine smoking gun, the fossilized dinosaur heart only gives rise to new local underdetermination problems.

8. The Roles of Background Theories in Historical vs. Experimental

https://doi.org/10.1086/426851 Published online by Cambridge University Press

Science. Permanent local underdetermination problems are widespread in historical science, but less common in experimental science. Why? The main reason for this has to do with the different roles that background theories play in historical vs. experimental science. In historical science, as I hope to have shown, background theories about information-destroying historical processes lead to widespread local underdetermination problems, because they mean that condition d in the above analysis of such problems will very often be satisfied. Such background theories imply that there are a great many things that scientists will never, ever know about the distant past. Or to put it another way, those background theories serve (or should serve) as a check to the epistemic ambitions of historical researchers. When historical scientists go looking for a smoking gun, they are, to a large degree, at nature's mercy. Although they can develop new technologies for identifying and studying potential smoking guns, such as the CT scans used by Fisher et al. to study the internal structure of the alleged dinosaur heart, historical scientists can never manufacture a smoking gun. If, in fact, every single dinosaur heart was destroyed by the fossilization process, there is nothing anyone can do about it.

On the other hand, background theories play a very different role in experimental science. Whereas background theories about information-destroying processes must limit the epistemic ambitions of historical scientists, a different set of background theories serves (and should serve) to enlarge the epistemic ambitions of experimentalists. Scientific realists, such as Richard Boyd (1985), have long emphasized the dialectical relationship between theory and method in experimental science. Experimental design always depends heavily upon background theories that tell scientists how to build experimental apparatus that will enable them to manipulate certain test conditions. This experimental manipulation then gives them a way to test new theories and hypotheses that, if confirmed, may provide new clues for the design of future experiments. In experimental science, background theories serve as guides for the design of new experiments whose purpose is to produce results that evidentially discriminate between weakly empirically equivalent hypotheses. A good example of this is Ian Hacking's (1983, chapter 16) discussion of the design of the polarizing electron gun ("PEGGY II"). Scientists and engineers relied on a variety of background theories and assumptions when designing PEGGY II—about the properties of gallium arsenide crystals, about the behavior of lasers, and so on. Then they used the device to test further hypotheses about microphysical entities (more specifically, to determine whether there are parity violations in weak neutral interactions). One lesson to be learned from such cases is that background theories enable experimentalists to build new apparatus, which in turn enable them to produce new phenomena—new evidence—in the lab.

https://doi.org/10.1086/426851 Published online by Cambridge University Press

Here, then, is the central argument of this paper:

P1. In prototypical historical science, background theories tell us how historical processes destroy information. Background theories do not usually play this role in experimental science.
P2. In classical experimental science, by contrast, one of the main functions of background theories is to serve as guides for the design of new experimental apparatus whose purpose is to produce new evidence that breaks evidential ties between weakly empirically equivalent hypotheses. Background theories do not usually play this role in historical science.
C1. Hence, there is at least one good reason for thinking that local underdetermination problems will be more widespread in historical than in experimental science.
C2. Hence, prototypical historical science is, in one sense, epistemically inferior to classical experimental science.

9. Conclusion. In this paper, I have identified two problems with Cleland's argument for the thesis that prototypical historical science is, epistemically speaking, just as good as classical experimental science. The first problem is that since overdetermination of events by their later traces (in Lewis's sense) is perfectly compatible with permanent local epistemic underdetermination, no interesting epistemological conclusions follow from Lewis's thesis of the time asymmetry of overdetermination. The time asymmetry of overdetermination does not lend any support to Cleland's thesis concerning the relative epistemic status of historical vs. experimental science. The second problem is that there is at least one independent reason for thinking that Cleland's conclusion is false—i.e., that historical science is, in one respect, epistemically inferior to experimental science. The different roles played by background theories in historical as opposed to experimental science give us some reason for thinking that permanent local underdetermination problems will be more common in the former than in the latter.

I do not regard this issue as completely settled. There may be other considerations that I have not dealt with here which should lead us to think that historical science is better off (or experimental science worse off), epistemically speaking, than I have suggested. Although many people seem to share the view experimental methods confer some epistemic advantages that are not to be had in historical science, it is hard to find any good arguments for this view. I hope to have shown that one way to justify this view is to stress the different roles played by background theories in historical vs. experimental science. In historical science, the

https://doi.org/10.1086/426851 Published online by Cambridge University Press

background theories tell us *how nature has destroyed the evidence*. In experimental science, they tell us *how to make new evidence*.

It is worth emphasizing, in closing, that I do not mean to say that any particular historical theory, such as Darwin's evolutionary theory, is less well confirmed than any particular theory about tiny things, such as quantum theory; or that prototypical historical science is in any way less scientific than classical experimental science; or that we do not really have scientific knowledge of the past; or that historical science is less worth doing than experimental science; or that historical science is in any way less rational or less objective than experimental science. All that I claim to have shown is that local underdetermination is a somewhat bigger and more pervasive problem in historical science than in experimental science, and that means that there is one sense in which the former is epistemically inferior to the latter, after all.

Cleland's most important insights are actually compatible with this modest result. For example, she justly infers, from the fact that historical scientists and experimental scientists exploit different aspects of the asymmetry of overdetermination, that "neither practice may be held up as more objective or rational than the other" (2002, 476).

REFERENCES

Alvarez, Luis W., et al. (1980), "Extraterrestrial Cause for the Cretaceous-Tertiary Extinction", *Science* 208: 1095–1108.

Barlow, Connie (2000), *The Ghosts of Evolution: Nonsensical Fruit, Missing Partners, and Other Ecological Anachronisms*. New York: Basic Books.

Boyd, Richard (1985), "Lex orandi est lex credendi", in Paul M. Churchland and Clifford A. Hooker (eds.), *Images of Science*. Chicago: University of Chicago Press, 3–34.

Cleal, Christopher J., and Barry A. Thomas (1999), *Plant Fossils: The History of Land Vegetation*. Rochester, NY: The Boydell Press.

Cleland, Carol E. (2002), "Methodological and Epistemic Differences between Historical Science and Experimental Science", *Philosophy of Science* 69 (3): 474–496.

Evans, David A. D. (2000), "Stratigraphic, Geochronological, and Paleomagnetic Constraints upon the Neoproterozoic Climate Paradox", *American Journal of Science* 300: 347–433.

Fisher, Paul E., et al. (2000), "Cardiovascular Evidence for an Intermediate or Higher Metabolic Rate in an Ornithischian Dinosaur", *Science* 288: 503–505.

Hacking, Ian (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hoffman, Paul, et al. (1998), "A Neoproterozoic Snowball Earth", *Science* 281 (5381): 1342–1346.

Janzen, Daniel H., and Paul S. Martin (1982), "Neotropical Anachronisms: The Fruits the Gomphotheres Ate", *Science* 215: 19–27.

Jenkins, Gregory S. (2000), "The 'Snowball Earth' and Precambrian Climate", *Science* 288: 975–976.

Kemp, Thomas S. (1999), *Fossils and Evolution*. Oxford: Oxford University Press.

Kirschvink, Joseph L. (1992), "Late Proterozoic Low-Latitude Global Glaciation: The Snowball Earth", in J. William Schopf and Cornelis Klein (eds.), *The Proterozoic Biosphere: A Multidisciplinary Approach*. New York: Cambridge University Press, 51–52.

https://doi.org/10.1086/426851 Published online by Cambridge University Press

Kukla, André (1994), "Non-empirical Theoretical Virtues and the Argument from Underdetermination", *Erkenntnis* 41: 157–170.
——— (1996), "Does Every Theory Have Empirically Equivalent Rivals?", *Erkenntnis* 44 (2): 137–166.
Laudan, Larry, and Jarrett Leplin (1991), "Empirical Equivalence and Underdetermination", *Journal of Philosophy* 88: 269–285.
Lewis, David (1979), "Counterfactual Dependence and Time's Arrow", *Nous* 13: 455–476.
Norman, David (1988), *The Prehistoric World of the Dinosaur*. Greenwich, CT: Brompton Books Corporation.
Rowe, Timothy, E. F. McBride, and Paul C. Sereno (2001), "Dinosaur with a Heart of Stone", *Science* 291: 783a.
Sober, Elliott (1988), *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, MA: MIT Press.
Stanford, P. Kyle (2001), "Refusing the Devil's Bargain: What Kind of Underdetermination Should We Take Seriously?", *Philosophy of Science* 68 (Proceedings): S1–S12.
Thulborn, Richard A. (1990), *Dinosaur Tracks*. New York: Chapman and Hall.
Wilson, Jeffrey A., and Matthew T. Carrano (1999), "Titanosaurs and the Origin of 'Wide-Gauge' Trackways: A Biomechanical and Systematic Perspective on Sauropod Locomotion", *Paleobiology* 25 (2): 252–267.

https://doi.org/10.1086/426851 Published online by Cambridge University Press