Teaching Plan:
The Scientific Method for the Historical Sciences

Phil Andrews
philandrews@pre-historic.com

"Equipping students with the tools of historical reasoning is not only an educational priority, it is a democratic imperative." (Ron Gray, 2025)

Summary

The historical sciences are distinct from predictive/experimental sciences because:

1. We cannot go back in time and observe the proposed events and generally we cannot replicate the proposed events in the present.
2. The historical sciences, in their purist form, do not make predictions about the future and this has two key outcomes:
    a. the historical sciences do not discover new regularities of nature; and
    b. narratives about the deep pre-historic past can be completely wrong and still do not pose any future safety risk to humans (although human history accounts and especially scientific test results do need to be correct).

Therefore, we need a scientific method specifically for the historical sciences that contains a feedback loop (like the feedback loop in the predictive/experimental scientific method) that enables increased reliability. A clear user-friendly version of the scientific method for the historical sciences contains at least two essential elements:

1. Researchers writing narratives about pre-historic events must show that the narratives only use and are theoretically consistent with the established regularities of nature.
2. Researchers must theoretically show that their narrative chain of events result in the world we see around us.

Background

The scholarly discussion of the philosophy of the historical sciences and its distinct methodology is a recent phenomenon in Philosophy of Science. Cleland was one of the first to widely promote and engage the discussion of the philosophy of the historical sciences and how they were distinct when she said (2001, 987): "In contrast, historical research involves explaining observable phenomena in terms of unobservable causes that cannot be fully replicated in a laboratory setting."

Jeffares followed up with his doctoral dissertation on this topic and published a paper (2008, 470), within which he observed: "The historical sciences then have two difficulties to overcome. They can't directly confirm their hypotheses about the past with observations due to the lack of access to the past. They can't confirm their hypotheses with contemporary observations because they are unique hypotheses about particular times or places. The result is a problem of confirmation. With no ability to observe their objects of enquiry, to repeat observations, or to intervene in processes, there is seemingly no way directly to confirm hypotheses. What I want to show in this paper is that… the historical sciences do utilise regularities… in overcoming these problems… the historical sciences seek regularities in the world, and have to do so in order to secure their claims about the past."

Q: What is a regularity of nature?

Q: What is an established regularity of nature?

Cleland (2011, 552) reasoned that the historical sciences are different to experimental science because they target long-past, token events, upon which controlled experiments cannot be conducted. Forber and Griffith (2011, 2) point out that we are not able to "reproduce or observe repetitions of most historical events… unlike testing of regularities." And Currie (2017, 929) shared his agreement with this when he said: "bygone eras are often beyond the reach of repeatable experiments."

In a nutshell, the historical sciences use present artifacts and the established regularities of nature to infer past events. These points of distinction cause the need for a specific scientific method for the historical sciences.

Experimental Limitations

Experimental science makes predictions that can be tested by future natural observations or in controlled lab experiments. Since we cannot go back in time to observe pre-historic events occur it would be nice if we could repeat the events in an experiment. However, as Jeffares (2008, 470) points out: "The historical sciences lack the ability to intervene in events in the same way as the experimental sciences for a further reason—that of scale. Repeating processes that take hundreds of years, or duplicating [large] processes like uplift of tectonic plates, is simply undoable. While some historical events are not at such a scale, many are, particularly in geology where both physical and temporal scale matters." Currie (2018, 229) agrees saying:

"Historical targets tend to be at large spatial and temporal scales – not easily admitting of experimental treatments."

To illustrate, think of the series of pre-historic events that led to the formation of the first star. Those events are unique events, because once the first star exists the following stars will never be the first one. And if following stars are near the first, then it influences them, presenting another set of unique events. This is also different to what we observe today with trillions of stars affecting each other. Star formation is one of many events with the problem of large spatial scale. Other pre-historic events have large temporal scale, for example biological evolution, as Popper (1978, 32) said: "natural selection, is difficult to test... really severe tests of the theory of natural selection are hard to come by, much more so than tests of otherwise comparable theories in physics or chemistry."

Q: What is the difference between a historical narrative and a prediction about the future?

Q: Can you think of an example of a deep past historical event that can not be repeated in an experiment?

Limits on Proposing New Regularities of Nature

The scientific revolutionaries showed us that observation and experimentation are key to discovering and establishing the regularities of nature. Therefore, since we do not have direct observations of the deep past and cannot directly experiment on the deep past, then we cannot use the deep past to discover new 'laws'/regularities of nature. Jeffares (2008, 470) explains: "The historical project is to come to conclusions about general processes; the regularities of the experimental sciences. This is both a means, necessary to secure claims about the past, and an end in itself… The historical sciences use background theories—theories about regularities—to secure the relationship between a hypothesis and a past cause… the historical sciences seek regularities in the world, and have to do so in order to secure their claims about the past."

Cleland (2002, 481) agrees: "But Wegener's hypothesis had a major defect. There was no known causal mechanism for horizontal continental motion." In other words, if a historical narrative is not based on a known mechanism/regularity of nature then the narrative/theory is defective. She continued with: "The basic idea behind narrative explanation is to construct a story—a coherent, intuitively continuous, causal sequence of events centering on a precipitating event and culminating in the phenomena (traces) in need of explanation."

If a historical scientist wishes their historical narrative to be described as coherent and intuitively continuous, then a key feature will be that the narrative aligns with the known/established regularities of nature, otherwise the narrative could be called fanciful of defective.

What we see here is that there are two essential elements of the historical sciences method:

1. Researchers must analyse their narratives that propose pre-historic events to show that the narrative is theoretically consistent with the established regularities of nature.

2. Researchers must theoretically show that their narrative chain of events result in the world we see around us.

Grim et al (2013, 2369) succinctly show emphasis on both: "In cases in which simulation is used for retrodiction, it is the input conditions that are read for new information. If the output conditions correspond to the current state of the world, and if the simulation's mechanism plausibly corresponds to ways in which we know the world to work, the input conditions indicate a possible previous state of the world." They clearly show how both aspects work alongside each other to achieve a reliable historical narrative. Thus researchers should describe/infer the past by using the established regularities of nature to show how initial conditions cause a series of events that result in the formation of the present world around us, including artifacts like fossils.

From the above assessment of what has been written about the laws/regularities of nature, it can be seen that the historical sciences only use established regularities of nature to construct narratives (like a type of applied science) and thus the historical sciences do not propose new regularities of nature. In contrast, the experimental sciences use test outcomes to reveal new, not yet noticed, regularities of nature.

Q: If a historical narrative includes a proposal of a regularity of nature that has never been observed in the present day, why would it be an unreliable narrative?

Verifying Historical Hypotheses: A Methodology

Firstly, it is important to note that even though historical science is distinct, it must continue to uphold the pre-eminence of the experimental sciences. Because historical science is a type of applied science it is completely dependent on experimental science defining what established regularities of nature are available for application (just like in engineering, medicine etc…). In other words, historical science cannot go off on its own tangent. It is completely subject to and limited by what is established by the experimental sciences via the traditional scientific method of hypothesis-experiment-results-revision. This recognition of the authority of the established regularities of nature over historical science cannot be emphasised enough.

It is well established that a key characteristic of science is that it concerns the regularities of nature. These regularities, once we have established reliable descriptions of them, enable us to use them to reliably predict the future in the context of applied science in fields like engineering, medicine etc... Leading up to this reliability have been centuries of experimentation. Generally, a proposed description of a regularity of nature made by the scientific revolutionaries came with a proposed prediction of the future outcome of a particular experiment or observation. The accuracy of the prediction was then assessed by comparing the actual outcome to the prediction and then the description/theory was adjusted/rewritten accordingly. This feedback loop process (Summers (1998), Justus (2005), Blachowicz (2009), Elliott (2012)) enabled descriptions to arrive at the level of extreme reliability we enjoy today. These are what are often called the established regularities of nature. It is the utilisation of a feedback process that increases the reliability of science, and this leads to Kuhn's (1996, 65) "precision of the observation-theory match."

Some theories, of course, are still in the experimental phase and these theories are not yet applied in engineering, medicine etc. They are not yet 'established' and pose too high a risk of causing harm to people. This is generally why applied science must only use established regularities of nature so that the predictions/applications are reliable and reduce the risk of causing harm to people.

Q: What might happen if Engineers or Doctors used unestablished theories?

The key then is that reliable descriptions of the regularities of nature are established by utilizing the feedback loop process of the traditional scientific method and iteratively moving towards reliability. Any proposed scientific method for the historical sciences must then also exhibit this feedback loop feature. Cleland's (2011, 567) "coherent, intuitively continuous, causal sequence of events" means the narrative must *only* include mechanisms that plausibly correspond to ways in which we know the world to work. Then a theoretical analysis of the narrative can reveal any violations of the established regularities of nature, or irregularities. Then the narrative can be revised as needed with a new proposed chain of events which can then also be analysed. This process is indeed a feedback loop, which means this method satisfies the requirement for such a process and will enable iterations to increase reliability of pre-historic narratives.

The second aspect of the Jeffares/Cleland method (which needs equal emphasis) is analysis of the 'predictions' of the model/narrative regarding what we should observe around us today. Jeffares (2008, 472) describes it this way: "Foreground theories make quite explicit claims; they effectively make predictions about what we should see in the record of the past." And Cleland agreed (2011, 567) that narratives should "culminate in the phenomena (traces) in need of explanation." If these 'predictions' in pre-historic narratives do not align with what we observe around us, then the narrative can be said to be 'falsified' and then it can be revised until it does align. This process is also a feedback loop, which means this second essential element of the historical sciences method satisfies the need to be a feedback process.

Therefore, both of these elements of the method are similar to the feedback loop in experimental science and are what Currie (2018) referred to as "virtual experiments". Evans and Thébault (2020, 9) offer a similar analysis in the context of inaccessible phenomena (similar in the sense that unique pre-historic events are also inaccessible phenomena): "In principle, it is thus perfectly possible for theories regarding inaccessible phenomena to be taken to be well supported or established based upon a suitably externally validated experiment and (where necessary) inductive triangulation strategy."

So, in the context of narratives that propose pre-historic events the externally validated 'experiment' is the mathematical model/simulation if it is used as a feedback loop. Similarly, analysis of a simple narrative using the two feedback methods can also be seen as an externally validated 'virtual experiment'.

Q: What are the two elements of the historical sciences method and how do they contain a feedback loop for iterative improvements?

Reliability and Risk

Karl Popper (1959, 316) famously said: "In so far as a scientific statement speaks about reality, it must be falsifiable." This has been debated by philosophers of science ever since, but the key is attempting to progress towards reliability. The historical sciences have that goal too. So, if a narrative about pre-historic chains of events is theoretically found to have events that do not align with the established regularities of nature, then the narrative can be said to be 'falsified'. With this feedback, the narrative is adjusted and re-analysed. In this way the feedback helps scientists to move towards more reliable theories.

Experiments are usually conducted with the intent to increase the accuracy and reliability of our descriptions of the regularities of nature. A key reason for this is that the established regularities of nature are used to improve the quality of human life and activity, so they must be reliable enough to be safely applied without causing harm.

Generally, societies have universal societal laws that require us to do no harm, so it is also the expectation that applied science also avoids causing harm to the highest standard possible. A scientific prediction about the future for example in engineering, medicine etc…, if it is wrong, can cause harm to people. How does science avoid causing harm? Applied predictive science uses only the established regularities of nature. Similarly, experiments on unestablished theories are conducted under strict safety protocols.

With the concept of safety risk in mind we can now review the risk environment of narratives that propose pre-historic events. The narratives do not make predictions about the future; therefore, they do not carry any risk for humans in the future. This is because these narratives are not used in Engineering/Medicine etc. Interestingly this means that narratives of pre-historic events can be completely wrong and still will not pose any risk to humans in the present or the future because applied science is never based on historical narratives about the deep pre-historic past. (This is different to human history accounts and scientific test result reports, which must be correct because they are applied in Engineering, Medicine etc.)

Regarding inductive risk, Douglas (2000), narratives that propose unique pre-historic events carry the highest inductive risk of all sciences because we cannot observe the deep past or experiment on it. But inductive risk is an indicator of how likely an inference is wrong. So narratives proposing pre-historic events have high inductive risk and low risk of causing harm to human. An example of the high inductive risk is noted by Odenwald (2022, 23) in his summary of the state of play regarding the origins of the universe when he concluded: "Of course, this entire story is highly speculative, even fanciful. It is based on theories or pieces of theories that remain largely unproven - or perhaps, one shudders to think, even unprovable." Odenwald openly admits the extremely high level of inductive risk by using the words "speculative", "fanciful" and "unprovable".

Thankfully we humans are very clever at contextualising and quickly judge that if a narrative about pre-historic events is speculative, it still does not endanger us. And so, like we do in all areas of human life, if there are no severe consequences or safety risk, we do not impose special safety protocols. So, the situation is that narratives that propose pre-historic events can be completely wrong (due to high inductive risk) without consequence, that is, very low safety risk to human life. Awareness of this situation will help motivate researchers to employ the scientific method for the historical sciences.

A case study:

Scott et al's (2000, p 2) *The Burgess Shale*: "Water caused sediments to flow in the form of mudslides, quickly burying the living organism in moving sediment. Fossils are therefore found in random orientation, indicative of a violent mudslide engulfing many of the Cambrian organisms. There is also evidence that these organisms died instantly. First, in the presence of an anaerobic environment, such as mud, marine invertebrates normally curl up upon dying. Fossils of the Burgess Shale locality do not exhibit this coiling. Secondly, there is no evidence of any attempt by these organisms to burrow out of their mud "prison." Killed instantly by the mudslides, preservation began immediately."

Q: How is present information used to infer the past events?

Q: How well do the retrodicted chain of events align with the established regularities of nature?

Q: Do the narratives chain of events result in what we observe today?

Q: Do you conclude that this retrodictive narrative is reliable or unreliable?

Conclusion

The historical sciences are distinct because we cannot go back in time and observe the proposed events and generally, we cannot replicate the events in the present.

Furthermore, the historical sciences, in their purist form, do not make predictions about the future and this has two key outcomes; one, the historical sciences do not discover new regularities of nature and two, narratives about the pre-historic past can be completely wrong and still do not pose any future safety risk to humans. Because of these distinctions the historical sciences must have their own scientific method to enable narratives to be reliable. This user-friendly method has two necessary elements that allow the historical sciences to be reliable. One; researchers writing narratives about pre-historic events must show that the narratives only use and are theoretically consistent with the established regularities of nature and two; researchers must theoretically show that their narrative 'predicts' the world we see around us. This method enables a feedback loop process, like we already have in the experimental/predictive sciences, and this allows historical scientists to iteratively improve their narratives and increase reliability of the historical sciences.

A note on nomenclature

It is possible that referring to the historical sciences as Retrodictive Science might be a clearer way moving forward. Jeffares (2008, 470) says: "The historical sciences seem to make retro-dictions— claims about the past." The Oxford Dictionary of Philosophy (2024) defines Retrodiction as: "The hypothesis that some event happened in the past, as opposed to the prediction that an event will happen in the future." And Merriam-Webster (2024) define Retrodiction as: using present data to infer past events.

References

Betini, Gustavo, Tal Avgar and John Fryxell. 2016. "Why Are We Not Evaluating Multiple Competing Hypotheses in Ecology and Evolution?" *Royal Society Open Science*, 5 December 2016.

Blachowicz, James. 2009. "How Science Textbooks Treat Scientific Method: A Philosopher's Perspective" *The British Journal for the Philosophy of Science* 60:2 (https://www.journals.uchicago.edu/doi/abs/10.1093/bjps/axp011)

Brustein, Ram, and Judy Kupferman. 2012. "The Creation of the World-According to Science." *History And Philosophy Of The Life Sciences*: 1202.0623. (https://arxiv.org/abs/1202.0623).

Chapman, Robert , and Alison Wylie. 2016 Evidential Reasoning in Archaeology. London: Bloomsbury Academic, 2016. Debates in Archaeology. Debates in Archaeology. Bloomsbury Collections. (http://dx.doi.org/10.5040/9781474219167).

Cleland, Carol. 2001. "Historical science, experimental science, and the scientific method." *Geology*, (2001) 29 (11): 987–990. https://doi.org/10.1130/0091-7613(2001)029<0987:HSESAT>2.0.CO;2

Cleland, Carol. 2002. "Methodological and Epistemic Differences between Historical Science and Experimental Science." *Philosophy of Science*, 69(3) 2002, 474-496. https://doi.org/10.1086/342455

Cleland, Carol. 2011. "Prediction and Explanation in Historical Natural Science.", *The British Journal for the Philosophy of Science*. 62:551–582 (DOI:10.1093/bjps/axq024).

Currie, Adrian. 2017. "Hot-Blooded Gluttons: Dependency, Coherence, and Method in the Historical Sciences." *The British Journal for the Philosophy of Science* 68:4, 929-952 (https://www.journals.uchicago.edu/doi/10.1093/bjps/axw005).

Currie, Adrian. 2018. *Rock, Bone, and Ruin: An Optimist's Guide to the Historical Sciences*.

United Kingdom: MIT Press Limited.

Currie, Adrian. 2019. *Scientific Knowledge and the Deep Past: History Matters*. Cambridge University Press, 2019.

De Cruz, Helen. 2021. "Religion and Science." In *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), edited by Edward Zalta. (https://plato.stanford.edu/archives/win2021/entries/religion-science)

Descartes, René. 1644. *Principles of Philosophy.* Translated by Miller and Miller, Holland:Kluwer Academic Publishers, 1982. First published in Latin, in 1644, p. 3.

Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science,* Dec., 2000, Vol. 67, No. 4 pp. 559-579. (http://www.jstor.org/stable/188707)

Elliott, Kevin C. 2012. "Epistemic and methodological iteration in scientific research." *Studies in History and Philosophy of Science Part A* 43:2:376-382, June 2012. (https://doi.org/10.1016/j.shpsa.2011.12.034)

Evans, Peter, and Karim Thébault. 2020. "On the Limits of Experimental Knowledge." *Philosophical Transactions A Royal Society* 378: 20190235. (http://dx.doi.org/10.1098/rsta.2019.0235)

Forber, Patrick, and Eric Griffith. 2011. "Historical Reconstruction: Gaining Epistemic Access to the Deep Past." *Philosophy, Theory, and Practice in Biology 3* 2013:0604 (DOI 10.3998/ptb.6959004.0003.003.)

Gray, Ron E. 2025. "Reconstructing the Past Scientific Reasoning in the Historical Sciences Through the Lens of Endosymbiotic Theory." *Science & Education* 14 June 2025 (https://doi.org/10.1007/s11191-025-00664-x)

Grim, Patrick, Robert Rosenberger, Adam Rosenfeld, Brian Anderson and Robb Eason. 2013. "How Simulations Fail." *Synthese* 190:2367–2390. (https://doi.org/10.1007/s11229-011-9976-7)

Gunn, James. 2007. In Cho, Adrian. "A Singular Conundrum: How Odd Is Our Universe?" *Science* Vol. 317, Issue 5846, pp. 1848-1850 (DOI: 10.1126/science.317.5846.1848)

Jeffares, Ben. 2008. "Testing Times: Regularities in the Historical Sciences." *Studies in History and Philosophy of Biological and Biomedical Sciences*, 39 (2008) 469-475. (https://doi.org/10.1016/j.shpsc.2008.09.003)

Justus, James. 2005. "Qualitative Scientific Modeling and Loop Analysis." *Philosophy of Science* 72:5: Proceedings of the 2004 Biennial Meeting of The Philosophy of Science Association, December 2005, pp. 1272 – 1286 (DOI: https://doi.org/10.1086/508099)

Kuhn, Thomas. 1996. *The Structure of Scientific Revolutions 3rd Edition*, Chicago: University of Chicago Press. (First published 1962.)

Maudlin, Tim. 2012. Quoted by Ross Andersen in "What Happened Before the Big Bang? The New Philosophy of Cosmology." *The Atlantic*, 19 January 2012.

Mayr, Ernst. 2009. "Darwin's Influence on Modern Thought." *Scientific American:*20091124. (https://www.scientificamerican.com/article/darwins-influence-on-modern-thought1/).

McGrew, Timothy. 2019. "Miracles." In *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), edited by Edward Zalta. (https://plato.stanford.edu/archives/spr2019/entries/miracles).

Merriam-Webster. 2024. (https://www.merriam-webster.com/dictionary/retrodict).

Nelson, Richard. 2016. "The Sciences are Different and the Differences Matter." *Research Policy* 45:1692–1701. (https://www.sciencedirect.com/science/article/pii/S0048733316300981)

Odenwald, Sten. 2022. "Imagining Our Infant Universe." *Astronomy* April 2022.

Oxford Dictionary of Philosophy. 2024. (https://www.oxfordreference.com/display/10.1093/acref/9780199541430.001.0001/acref-9780199541430-e-2710)

Popper, Karl. 1959. *The Logic of Scientific Discovery.* London: Routledge, 2002, p. 316. (First published in 1959).

Popper, Karl. 1978. "Natural Selection and the Emergence of Mind." *Dialectica* 32:339-355. (http://www.jstor.org/stable/42970324)

Scott, Erin, Lara Kirkner, Jane Shin, Veeral Desai and James Chan. 2000. *The Burgess Shale*, Berkeley University of California, 5 May 2000.

Skell, Philip. 2005. "Why Do We Invoke Darwin?" *The Scientist*, 28 August 2005. (https://www.the-scientist.com/opinion-old/why-do-we-invoke-darwin-48438).

Smeenk, Christopher and George Ellis. 2017. "Philosophy of Cosmology." In *The Stanford Encyclopedia of Philosophy* (26 Sep 2017), edited by Edward Zalta. (https://plato.stanford.edu/archives/win2017/entries/cosmology/)

Summers, Richard L. 1998 "Computer Simulation Studies and the Scientific Method." *Journal of Applied Animal Welfare Science* Vol 1:2 119-131, 1998. (https://doi.org/10.1207/s15327604jaws0102_3)

Turner, Michael. 2009. Video of Panel 1 at the Origins Symposium, Arizona State University, 3 April 2009. (http://thesciencenetwork.org/programs/origins-symposium).

Wylie, Alison. (2020). Radiocarbon Dating in Archaeology: Triangulation and Traceability. In: Leonelli, S., Tempini, N. (eds) Data Journeys in the Sciences. Springer, Cham. (https://doi.org/10.1007/978-3-030-37177-7_15)