Evolutionary Anthropology 25:271–275 (2016)

Gang Warfare

What are the criteria for deriving assertions from facts in science?

KENNETH M. WEISS

On a bowered bench in our university Arboretum, surrounded by glorious blooms, bees, and butterflies sat a student, deeply absorbed in a book. Beside him were several other books, and among them I recognized ones I also have. Interrupting his concentration, I asked what he was studying. He said he's a Philosophy major, specializing in the philosophy of science.

The philosophy of science concerns the way we develop inferences about nature in the absence of mathematical proof. For many years, I've discussed this subject in courses and presentations, and found that even experienced professionals, including anthropologists, often have only informal ideas about the way science develops theory or makes inferences. They've often said that my presentation helped them think more critically about their field. So I thought I'd lay out some of the issues as I see them.

THE MANY FACES OF EVIDENCE

We've all been taught that the "scientific method" is the way we come to our conclusions about the nature of the world. We form a hypothesis and use it to make specific predictions. We then test the predictions by controlled observation or experiment; modify the hypothesis in light of the results; and repeat these steps until there are no longer discrepancies between results and predictions.

When or whether we actually stick to this scenario is debatable. For starters, in the real world there are always at least some discrepancies between results and predictions. More importantly, the scientific method doesn't tell you how to set up your test or interpret its results. This is often treated rather casually both in teaching and practice. Many useful references discuss various criteria that are applied to understanding nature.[1–13] However, these criteria weren't found inscribed on a tablet in the desert. We invented them, and each one has problems

1. Deduction from theory

Science is about causation, and if we have a theory about the factors and forces at work, it tells us what, under a given set of conditions, we should observe. Going back at least to Aristotle is the idea that we somehow can formulate ideas and means of reasoning (deduction) from assumptions to predict what will happen if some particular conditions apply. From this viewpoint, a theory can have intuitive or conceptual origin in first principles of nature, which will be confirmed by observation. Unfortunately, intuition varies. We don't all agree on the premises we adopt, upon which valid deductions rest.

2. Induction

Observation, not abstract reasoning, is the only true basis of understanding. If we're after nature's laws, then one-off observations aren't convincing. But if we consistently observe a result, we can believe it happens for a specific causal reason. A problem with this criterion is that it only takes only one contradiction to undermine what we had thought was a "law."

3. Replicability

Related to induction is the point that if we have made correct inferences from seeing some apparent cause and effect, and if our understanding is right, we should always be able, at least in principle, to repeat the same conditions and achieve identical results. Unfortunately, it's not easy to have the "same" conditions or even to know if we have, especially in the social and life sciences.

4. Predictability

A correct understanding of cause and effect should allow correct predictions in relevant new situations. If you don't obtain an expected result, you question your theory. Prediction is problematic for at least two reasons. One reason is that you can obtain a predicted outcome in the absence of your favorite cause if the outcome can have other causes; the other is that you could not get the predicted outcome if it depends on some unaccounted-for additional factors, including errors in your study or measurements.

5. Verifiability

A school of early twentieth-century thought known as the "Vienna Circle" (Wikipedia: Vienna Circle) held a view called logical positivism, which held that meaningful assertions about

Ken Weiss is Evan Pugh Professor Emeritus of Anthropology and Genetics at Penn State University. Email: kenweiss@psu.edu

DOI: 10.1002/evan.21509
Published online in Wiley Online Library (wileyonlinelibrary.com).

the real world, as opposed to subjective judgments about such topics as love or art, must be verifiable, at least in principle. You must be able to specify how your assertion could be tested. That was seen as a definition of causal objectivity in nature.

6. Falsifiability

Karl Popper (1902-1994), a philosopher of science, is known for leaving the Vienna Circle, in part because of the verifiability criterion (Wikipedia: Karl Popper). One can find evidence consistent with one's hypothesis, but this does not imply that the hypothesis is the reason for the evidence. We can't ever prove a theory is true, he held, but it must be falsifiable. We can falsify ideas by experiments that don't find a predicted outcome. Over time, we eliminate what's not true and, as Sherlock Holmes put it in *The Sign of the Four*, "Once you eliminate the impossible, whatever remains, no matter how improbable, must be the truth."[25: ch. 6]

Falsifiability is often invoked in what I think is a rather conveniently self-serving way when a research study provides support for its originator's idea; that is, when the idea was not falsified. However, it is notorious that negative results aren't published or that data can be presented in a way that supports rather than falsifies an hypothesis. But more profoundly, it's important to realize that falsifiability isn't the gold standard it's purported to be. Negative results can be misleading not because the hypothesis was wrong, but for other reasons the investigator isn't aware of. There can be various measurement, sampling, or other errors about the study itself that lead to inconsistency, even with a correct theory. For example, statistically "nonsignificant" results can be taken as evidence against a theory, rejecting a null hypothesis. However, such decision points are only probabilistic in nature and might just result from bad luck in choosing one's study sample.

7. Parsimony

The medieval philosopher William of Occam (1287-1347) expressed a criterion for constructing scientific theories, known as Occam's Razor (Wikipedia: William of Ockham; and Occam's razor). According to that criterion, you should prefer the simplest hypothesis that is consistent with the data. Nature is sleek in its laws, and we don't gain by making *ad hoc* subtheories or exceptions. This works pretty well in chemistry and physics; one would not need to stipulate a distribution of weights or charges among different electrons, because assuming they're all identical works well.

Unfortunately, Occam's Razor doesn't work for evolutionary and social biology because the fundamental nature of life is that it involves local, *ad hoc* context-dependent variation. If evolution is about anything, it's about not being the same: Mutations and chance affect the genetic variation present in any time and place, so no two contexts are identical. That is the central fact that enables evolution, as chance generates and natural selection screens variation. Every individual, indeed, every cell in every given individual, is genetically different. Parsimony is useful in some settings, and often glibly invoked to defend one's interpretation of findings, but it is not an ironclad standard for evolutionary, biological, or sociocultural inference. It is often not easy to decide which variables are worthy of including or ignoring, nor even what is the 'simplest' hypothesis! Evolution need not follow any direct much less 'simple' path.

8. Probabilistic causation

We tend to think of causation as deterministic; that is, under given conditions, given results will follow with certainty. But in practice, and especially in some areas of physics, causation seems to be inherently probabilistic. Under given conditions, result X and result Y will occur with fundamental law-like probabilities we can estimate. How nature determines the probability values is a deep question. In the life and social sciences, it's usually, if tacitly, assumed that the apparently probabilistic nature of a phenomenon arises because we simply don't know enough about its cause and that if we did, prediction could be precise. Flipping coins seems inherently probabilistic, but if we knew all the conditions it would become essentially deterministic.

Many aspects of life appear to be probabilistic, raising questions about how we can make assertions about them. The answer is that we use criteria like statistical significance, taking the statistical improbability of a result as a guide to when to accept a proposed cause as "true." But that's a subjective judgment,[14] a deep problem receiving prominent recent attention in the context of widespread nonreplicable or exaggerated findings of biomedical studies.[15–21] That judgment applies even more to sociocultural and even evolutionary science. One's result is often sample-dependent, and the causes of outcomes may vary among samples. For example, multiple genetic or other local, *ad hoc*, or context-dependent factors affect one's reproductive fitness.

9. Probabilistic theory itself

In some circles, it is acknowledged that we can never have perfect criteria for scientific inference. All we can hope for is improved confidence in a theory. Sometimes our advocacy of a point of view is simply subjective or intuitive, but at other times we can express confidence more formally. One way is to compare ideas we may have about what's going on and see how much more probable one idea is than the other. When you have specific, quantifiable alternatives in mind, the result provides formal statistical support for your idea. Is a variant in this gene found more often in those with some trait you care about? Is a male equally likely to have sons and daughters (transmit an X, or Y chromosome to a given offspring), or is there a reason why, for example, males would father fewer daughters than sons?

Sex ratio illustrates another way of improving confidence by revising the value of some key parameter as you accumulate data. You start out

expecting fathers to sire males and females with equal probability. But what if your study finds 55% males? Bayes' Theorem in probability theory (Wikipedia: Bayes' theorem), provides a way to use the new result to alter your accepted value of the sex ratio. The result could, of course, be a quirk of chance. You can never know what is really "true," but the approach can make you believe in a value greater than 50% (equal sex ratio) and lead you to ask why that may be. Could some aspect of X versus Y chromosomes give Y-bearing sperm an advantage? Is this universal or does it apply only to males with a particular genotype or does it occur only under some conditions we happen to have seen? Plausible reasons can sometimes be suggested, but it's tricky because they may not always apply among or even within populations. We can end up confidently asserting that the sex ratio is ½ … except when it isn't. That assertion that only sounds scientific.

Bayesian and likelihood approaches have fervent advocates who argue that other methods of inference only give illusions of proof. At least, they say, their approach allows us to adjust and refine what we think might be true, and that's all we can hope for.

10. Who cares? Use whatever you want!

In reality, here's the most often used criterion: If you do a study but your data don't really support your hypothesis, various sorts of manipulation or altering of acceptance criteria can still lead you to claim to have found what you wanted to find. This is subtle and never shows up in print. The above nine criteria for inference are invoked selectively, even if with great assertiveness, while other criteria are often carefully ignored in ways specific to a given paper.

One needn't be entirely cynical about this because, often, an important criterion, for various practical reasons, cannot be applied. After a lot of work and commitment to an idea, investigators are understandably determined to make a major publication of it. Still, causation can be complex, and data mining, eliminating outliers, excusing aberrant results as laboratory artifacts, choice of significance cutoff criteria, adjusted ways of explaining the "importance" even of a weak finding, and so on, are routinely part of scientific reports. Among other things, the statistical nature of most of the evidence doesn't help when it comes to issues like prediction and replicability.

Of course, with experience, one's intuition about a situation may be correct, so there are reasons to believe in your ideas, even if trying to prove them convincingly leads to these sorts of gerrymandering of the data.

In the end, there really are no formal self-imposed means of making scientific inferences. All ten criteria are in play, as are others of a more sociological nature, such as the need to compete for grants, tenure, publications, or to sell research supplies, and more. Scientific inference is more subjective than we like to think.

A FEW EXAMPLES

Genetic or evolutionary assertions are widespread and often presented as established truth. The following examples are not documented literature reviews, and I'll intentionally be very skeptical in order to show how assertions might demand closer attention to evidentiary criteria. What criteria would you use to convince yourself of these or similar assertions, reject them, or formulate a better alternative?

1. Does this gene "cause" that disease?

Alzheimer's disease (AD) is a common dementia that mainly affects older people in a devastating way. Many studies report that tangles of neurons called plaques cause AD, and that a genetic risk factor is the E4 variant of the ApoE gene (Wikipedia: Senile plaques; Neuropathology). E4 is produced by a combination of nucleotide variants in two sites along the ApoE. But does the association of E4 with AD mean it causes AD? The associated lifetime risk is far from 100%, and most people who have the variant remain unaffected. Plaques have been causally associated with other specific genes and are found during post mortem examinations of unaffected people, including ones who do not have the E4 variant. DNA sequencing has found other variation in or near this gene, some of which relates to expression of the gene rather than its coded protein structure. Also, different ethnic groups, particularly African-derived people, may have little or no excess risk associated with E4. Further, the E4 allele is the ancestral one in apes, which don't seem to get AD. Further, cholesterol levels, among other life-style factors, are associated with AD. Variants in two other genes called presinillin*s* seem to confer very high risk, but they're very rare in the population. (Also, the genes are misleadingly named, as if they evolved to cause AD). Is it these, or ApoE, or other genes and different environmental exposures that are responsible? If so, why is the risk so "statistical," with not everyone affected, and what, exactly, is the risk? How should our ten criteria be used to try to decide?

One of the clearest, most dangerous genetic risk factors for breast cancer is mutation in the BRCA1 gene. In some study cohorts, risk by the age of 60 years can approach 90%, and nobody doubts that this gene has a causal role. Yet in different study groups, in different places, or on different birth cohorts (for example, women before versus after 1940), the risk by the age of 60 years can differ by about two-fold. That is, risk is somehow sample-specific. BRCA1 appears to function by detecting and repairing DNA mutations in cells, removing otherwise cancer-causing changes. Such mutations aren't detected and fixed in persons with the mutant BRCA1 gene. Yet, curiously, while the gene is expressed in most or all tissues, only a few tissues show excess risk in those with the dangerous BRCA1 mutations. Clearly, variation in other genes, life-style factors, or something entirely unknown is responsible for who develops the disease.

These are among the clear cases. There's no doubt that they and many

others we see in the daily news are involved in their associated traits, but in ways so variable and unclear that their replicability and predictive power is problematic. Such facts aren't hidden from those who promise near-miracles of "precision" genomic medicine that rest on predictability, so that it's fair to ask by what criteria or in what sense do such genes "cause" their particular disease? I think even far more circumspection is called for regarding claims of predictable genetic causation of socially sensitive behavioral traits, even though genetic variation must contribute to them as it does to all traits.

2. Did human longevity evolve for child care reasons?

An idea that has been widely argued is that humans' dramatically extended life expectancy is due to a fitness advantage conferred on grandchildren who have a long-lived grandmother to care for them. A dowager grandma provides care and resources for her grandchildren and hence increases the fitness of the genotypes that led her to live long enough to be around when her grandchildren were born. Observing the grandmother effect in some populations is consistent with its having a genetic basis favored by natural selection, though we should be able to find specific genotypes that are responsible and show that they are the reason for the grandmother's longer life. Given how hard it is to find genes "for" late-onset diseases even with massively large studies, it would be daunting to collect data adequate to prove the evolutionary hypothesis based on small local populations, much less to generalize it or extrapolate the effect over the past million years of hominid evolution. Indeed, in some societies the elderly are a resource burden on their children and grandchildren.

I confess skepticism because, in my view, during our evolutionary past only a small fraction of humans lived to postreproductive grandparenting ages.[22] Moreover, their excess genetic correlation with their grandchildren is relatively weak relative to that of others in their typically small aboriginal bands in which everyone was closely related. Indeed, outside of modern protective environments, not only was late-age survival rare, but human life spans are not very different from what would be expected among mammals for our brain and body size and metabolic rate.[23,24] So, from the parsimony perspective, there's no need for a specific evolutionary explanation. That, of course, doesn't falsify the explanation: Nature, especially living nature, need not be parsimonious.

Still, I personally think mutual care-taking in hunter-gatherer groups is socially based rather than based on genetic evolution for longevity for that purpose. You don't have to agree. Fair enough, but then how would you prove the evolutionary assertion? That, after all, means finding specific genetic changes that are responsible and exist for that reason. Finding a grandparent effect in one population doesn't mean it would be found elsewhere, but that failure could be the result of other factors — replicability and induction are nice, but not essential for accepting a theory. In other words, is the grandmother scenario a nice observational local truth, a convenient just-so evolutionary story, or a real example of Darwinian evolution? What criteria would allow us to know?

3. Have game-playing tactics evolved genetically?

In a famous example of game strategy, two gang members are imprisoned and isolated from each other. The prosecutor privately offers release to each prisoner if he will betray the other. If they both betray, each serves 2 years in jail. If A betrays B, A goes free but B will have to serve 3 years. If both of them remain silent, both will have to serve 1 year. What do they do?

The various subtleties of the game have been written about extensively (Wikipedia: Prisoner's dilemma). Applications of this and other aspects of game theory address how competitive interactions "should" go if a simple model of evolution is true. Why would anyone cooperate? Why not cheat and get away with it? It is debatable whether humans or other animals actually follow what a Darwinian game theory predicts would evolve. But let's suppose that in some sense individuals do what a computer modeling of the dilemma says is best. Is this "genetic" in terms of its evolution? Can you show that it systematically affects reproductive fitness in a genetically specific sense? What sort of data, and how much of it, would you need to make the case? Again, as noted earlier, it takes massive samples to find genes associated with complex traits. In addition, genes almost always serve many functions simultaneously and their variants typically have very weak effect. So finding the variants is by no means the same as showing they're here because of successful ancestral strategizing.

Even if you found a fitness effect in some huge sample of people today, how could you show that playing the game had evolved by natural selection and was genetic for that reason? How could you exclude the plausible idea that we have evolved brains that are generally programmed to evaluate situations and come to conclusions rather than being preprogrammed for each situation we might encounter? Indeed, different people respond differently to given situations. Doing so clearly need not be "genetic." So did playing games like prisoner's dilemma evolve in a specific genetic way or did we evolve to be solvers of unanticipated problems?

We love nice, tidy evolutionary or genetic explanations for complex phenomena. The Darwinian drug suggests that almost every trait can be explained in terms of tractably specific genotypes, driven by natural selection into our nature in ways we can divine or even modify if we so choose. It leads to comforting explanations of an otherwise perplexing world. But when is that real and, more importantly for science rather than lore, how can we understand it?

HOW WE REALLY DECIDE: GANG WARFARE!

Science is populated by humans, and that may imply a certain

amount of tribalism. We use inferential criteria selectively, in what often is combat between groups of advocates of different points of view. Sometimes this is justified because there is no way to apply a particular criterion. But the rules of scientific combat are human-made and not given by any on-high Objectivity.

Advocates cite inferential criteria that favor their view based on the kind of data they've been able to assemble. Typically, the thrust of papers, even if filigreed at the end with careful caveats or with potentially undermining details buried in online "supplemental information," is to advocate a theory, viewpoint, or the importance of a finding.

We're trained by our academic gurus or make our name by our particular ideas or methods. Our students often offer loyalty to ideas we, after all, trained them in. We form alliances to set up research groups and have friendly reviewers for our papers and grant applications. Ideas, like misery, love company, and competing alliances are a part of the natural ecology of our social species. We like to have explanatory reasons for our alliances, so we pick what suits as we compete against other alliances.

Science is a soft form of gang warfare. We may feel we're not trained to think that way. However, academic advisors make it pretty clear to their colleagues and students that they must find reasons to defend their ideas, lest some competitors' ideas take the field. Criteria for inference are useful and powerful when they truly apply. But conveniently selective application is part of the gang's behavior. Each side has evidence to invoke; each has weaknesses and unknowns. The prevailing view at any time is what the dominant gang declares it to be. This has been called a "paradigm" by Thomas Kuhn[1] and a "thought collective" by philosopher of science Ludwik Fleck.[7,8] Ultimate proof is an elusive concept at best. Often, if not typically, consensus isn't driven by the rigorous application of inferential criteria that we like to think we follow.

All but the most skeptical of us accept that there are facts in the world that really do exist and that seem to be quite orderly. For several centuries in Western thought, we have referred to this regularity as the "laws" of nature. Even with distant telemetry, every place in the observable universe appears to follow the same laws of physics and new life forms we find show common genetic ancestry with the rest of life. Only the religious attribute natural laws to a personal God who dreamed them up and created a fully obedient universe. There may be other universes; there may be life on other planets that has no DNA basis, but there is every reason to believe that these would be chemical details relative to the overall concept of an orderly law-like existence. Why that's so is for philosophers to discuss, but our job as scientists and scholars is to go with what we see and try to identify nature's realities.

We have to explore by observation rather than finding law-books in our local library. Nothing has been handed down to us except the thoughts and experiences of millennia of our predecessors. Nature may be an accessible book, but not an open one. We have to wrestle with the facts we can observe, as we choose or are able to observe them, and try to establish consensus about what the laws are.

The result should be about nature, not the personalities holding professors' jobs. Much of it is. We strive to maintain objectivity in the process of deciphering the laws of nature and hope that we have arrived at proof. But in the reality of human life, the objectivity is often more claim than fact, clouded by subjective interests and our own empirical and intellectual frailties. Wrangling about how nature is, as members of advocacy gangs, is the nature of our jobs.

I was inspired to write this by encountering the young student in our Arboretum. He wants to understand how science works and to think about these issues professionally. There's an important role for people like him to play.

NOTE

[1]I welcome comments on this column: kenweiss@psu.edu. I co-author a blog on many of these topics at *ecodevoevo.blogspot.com*. I thank, for carefully and critically reading this manuscript, Anne Buchanan, Holly Dunsworth (twice!), and John Fleagle.

REFERENCES

1 Kuhn TS. 1970. The structure of scientific revolutions. Chicago: University of Chicago Press. xii.

2 Chalmers AF. 1999. What is this thing called science? Indianapolis: Hackett. xxii.

3 Howson C, Urbach P. 1993. Scientific reasoning: the Bayesian approach. La Salle, IL: Open Court.

4 Rosenberg A. 2000. Philosophy of science: a contemporary introduction. London: Routledge.

5 Quine WV. 1953. Two dogmas of empiricism. Philos Rev 60:20–43.

6 Hacking I. 1990. The taming of chance. Cambridge, UK: Cambridge University Press.

7 Fleck L. 1935. Genesis and development of a scientific fact (1976 reprint, with editorial comment by Kuhn and others). Chicago: University of Chicago Press.

8 Weiss KM. 2003. Ludwik Fleck and the art-of-fact. Evol Anthropol 12:168–172.

9 Jaynes ET, Bretthorst GL. 2003. Probability theory: the logic of science. Cambridge, UK: Cambridge University Press. xxix.

10 Hull DL. 2009. Darwin's science and Victorian philosophy of science. In: Hodge J, Radick G, editors. The Cambridge companion to Darwin, 2nd ed. Cambridge, UK: Cambridge University Press. p 168–192.

11 Stanford P. 2006. Exceeding our grasp. Oxford, UK: Oxford University Press.

12 Godfrey-Smith P. 2003. Theory and reality: an introduction to the philosophy of science. Chicago: University of Chicago Press.

13 Popper K. 2002. The logic of scientific discovery. London: Routledge.

14 Weiss KM. 2011. The 5% solution. Evol Anthropol 20:81–84.

15 Ioannidis JP, Ntzani EE, Trikalinos TA, et al. 2001. Replication validity of genetic association studies. Nat Genet 29:306–309.

16 Ioannidis JP. 2016. Why most clinical research is not useful. PLoS Med 13:e1002049.

17 Ioannidis JP. 2014. How to make more published research true. PLoS Med 11:e1001747.

18 Young NS, Ioannidis JP, Al-Ubaydli O. 2008. Why current publication practices may distort science. PLoS Med 5:e201.

19 Wasserstein RL, Lazar NA. 2016. The ASA's statement on *p*-values: context, process, and purpose. Am Stat 70:129–133.

20 Ioannidis J. 2016. Evidence-based medicine has been hijacked: a report to David Sackett. J Clin Epidemiol 73:82–86.

21 Goodman SN. 2016. Aligning statistical and scientific reasoning. Science 352:1180–1181.

22 Weiss KM. 1973. Demographic models for anthropology. Am Antiquity 38:1–186.

23 Weiss KM. 1989. Are the known chronic diseases related to the human lifespan and its evolution? Am J Hum Biol 1:307–319.

24 Weiss KM. 1981. Evolutionary perspectives on human aging. In: Amoss P, Harrell S, editors. Other ways of growing old. Stanford, CA: Stanford University Press. p 25–52.

25 Doyle AC. 1890. The sign of the four. Wikisource: https://en.wikisource.org/wiki/The_Sign_of_the_Four