Revisiting Chamberlin: Multiple Working Hypotheses for the 21st Century

1 author:

Louis P. Elliott
Northern Territory Government, Australia

7 PUBLICATIONS   **158** CITATIONS

Some of the authors of this publication are also working on these related projects:

Darwin Smoke Project 2005 View project

Mimosa Biocontrol View project

Revisiting Chamberlin: Multiple Working Hypotheses for the 21st Century

LOUIS P. ELLIOTT AND BARRY W. BROOK

The method of multiple working hypotheses, developed by the 19th-century geologist T. C. Chamberlin, is an important philosophical contribution to the domain of hypothesis construction in science. Indeed, the concept is particularly pertinent to recent debate over the relative merits of two different statistical paradigms: null hypothesis testing and model selection. The theoretical foundations of model selection are often poorly understood by practitioners of null hypothesis testing, and even many proponents of Chamberlin's method may not fully appreciate its historical basis. We contend that the core of Chamberlin's message, communicated over a century ago, has often been forgotten or misrepresented. Therefore, we revisit his ideas in light of modern developments. The original source has great value to contemporary ecology and many related disciplines, communicating thoughtful consideration of both complexity and causality and providing hard-earned wisdom applicable to this new age of uncertainty.

Keywords: Bayesian statistics, hypothesis testing, model selection, philosophy of science, statistical significance

For more than two decades, there has been sustained criticism of the appropriateness of using methods that rely solely on null-hypothesis testing for observational studies in science (e.g., Carver 1978, McBride et al. 1993, Anderson et al. 2000, Wade 2000, Johnson 2002). The disciplines of psychology, wildlife biology, and statistics have been in the forefront of this conflict between two qualitatively different inferential paradigms: model-selection methods, based on information theory, and null-hypothesis testing, based on a frequentist approach. But many other areas of biology and ecology have been implicated, including molecular biology, systematics, physical geography, medicine, and epidemiology (Johnson and Omland 2004). Perhaps this is because all these fields readily provide case studies in which multiple causative factors lead to real-world complexity that is difficult to reduce to a single, isolated mechanism.

Strong proponents of the model-selection paradigm have decried the use of null-hypothesis testing as outdated, and some have colorfully suggested that the practice of reporting *P* values should be "euthanized" on philosophical grounds (Anderson and Burnham 2002). Others have taken an equivocal stance, suggesting that the two inferential paradigms provide complementary tools for the investigator, and that hypothesis testing should be retained for manipulative experimental design (e.g., Johnson and Omland 2004). Stephens and colleagues (2005) proposed that it may be more profitable to distinguish between studies of univariate causality, in which null-hypothesis testing may be sufficient, and multivariate causality, in which model selection offers clear advantages (but see Lukacs et al. 2007).

Here we attempt to clarify some of the philosophical terrain relevant to this debate by discussing one of the key philosophical underpinnings of model selection. This is the concept of the method of multiple working hypotheses (MMWH), as described by the geologist T. C. Chamberlin in 1890, and later referred to by Platt (1964) in his notion of "strong inference." Although the term has become almost mainstream in ecology, we contend that the core meaning of Chamberlin's conceptualization has often been forgotten or misinterpreted over time, and that this needs rectification. For instance, a common mistake is to equate the MMWH with the method of developing alternative hypotheses. Yet systematic application of the latter method occurred at least as early as Francis Bacon (1620), whereas the former is qualitatively different in construction and was intended by Chamberlin to serve as a complement to the formal, "pure," or classic analytic method. Here we first describe the MMWH in general terms. Then we discuss its applicability to methodologies that not only allow (or require) the simultaneous appraisal of more than one hypothesis but explicitly accommodate various situations in which several hypotheses are simultaneously true.

Louis P. Elliott (e-mail: louis.elliott@nt.gov.au) was at the School for Environmental Research at Charles Darwin University in Australia when this essay was prepared; he is now with the Northern Territory Department of Natural Resources, Environment, and the Arts in Australia. Barry W. Brook (e-mail: barry.brook@adelaide.edu.au) is the director of the Research Institute for Climate Change and Sustainability, School of Earth and Environmental Sciences, University of Adelaide, South Australia 5005, Australia. © 2007 American Institute of Biological Sciences.

The method of multiple working hypotheses

The concept of the MMWH was advocated over a century ago by the geologist Thomas Chamberlin (1890) in a paper that was later reprinted in *Science*—a testament to the perceived importance of its content. "With this method," Chamberlin wrote, "the dangers of parental affection for a favorite theory can be circumvented" (Chamberlin 1890). Chamberlin's concerns have a timeless quality that makes his prose lucid and relevant even today. He contrasted the MMWH with the methods of the "ruling hypothesis" and the "single working hypothesis," and contended that the ruling hypothesis is the worse of the latter two. This is because investigators' affection or loyalty to a theory may lead them to collect evidence to support only the ruling theory, and not sufficiently consider alternative explanations. Chamberlin also criticized the single-working-hypothesis approach, said to be the method of the day: "Under the working hypothesis, the facts are sought for the purpose of ultimate induction and demonstration, the hypothesis being but a means for the more ready...arrangement and preservation of material for the final induction" (Chamberlin 1890).

The amendment that Chamberlin advocated is one familiar to all practitioners of science. However, like much cogent advice, it is easier to follow in theory than in practice: "[What is required] is to bring up into view every rational explanation of new phenomena, and to develop every tenable hypothesis respecting their cause and history" (Chamberlin 1890). This description approaches the true purpose of the MMWH: to circumvent the dangers of becoming emotionally attached to any given idea or hypothesis, and to work against the natural tendency to construct premature (or to require single and complete) explanations of phenomena. The approach is a carefully considered one, which poses such questions as "Is this really the full explanation?" or "Are we seeking to prematurely establish the truth of a single factor, when consideration of more than one may be more appropriate?"

Explicitly describing "synthetic cognition." Chamberlin claimed that after a period of time of following the application of the MMWH, a habit of thought develops that is analogous to the method itself:

> Phenomena appear to become capable of being viewed analytically and synthetically at once. It is not altogether unlike the study of a landscape, from which there comes into the mind myriads of lines of intelligence, which are received and coordinated simultaneously, producing a complex impression which is recorded and studied directly in its complexity. My description of this process is confessedly inadequate...but I address myself to naturalists who I think can respond to its verity from their own experience. (Chamberlin 1890)

This is a description of the processes by which researchers both tolerate and benefit from intellectual dissonance when confronted by complexity. Confounding variables and mechanisms can operate at different temporal and spatial scales, both in succession and simultaneously. This is often the case in diachronic problems in ecology, conservation biology, paleontology, epidemiology, medicine, geology, meteorology, and astronomy (Hilborn and Mangel 1997), in which it is often impossible to "wind back the clock" or to experiment on the systems involved; mechanisms must be inferred from other lines of evidence and later brought together into a consistent whole.

Chamberlin particularly stressed the importance of not being content with the idea of a single, often simple explanation, despite the pleasure that such an explanation may arouse in the mind of the researcher. Being a geologist, he used as a prime example the question of the origin of the Great Lakes basins. There are at least three "practically demonstrable" mechanisms by which the basins could have formed: (1) crust deformation, (2) preglacial erosion from rivers, and (3) glacial excavation (Chamberlin 1890). Whereas another researcher might have been content with one, or perhaps two, of these hypotheses, Chamberlin invoked all three of them, proposing that all three processes acted in temporal succession to produce the end result. This is commonly described as a "cascade" in ecology, and in medicine it corresponds to the distinction made between a primary and a secondary condition. For example, although a person may have died from heart disease, this illness would most likely have had prior contributing factors such as poor diet, lack of exercise, and smoking.

Sequential and simultaneous multiple working hypotheses.

Although Chamberlin did not make any formal distinction, it is useful to consider whether there may be different types of multiple working hypotheses. Causation, for instance, may occur as a series of sequential steps (figure 1a; e.g., a disease-ridden animal may be vulnerable to predation), or multiple factors (of varying importance) may operate simultaneously (figure 1b). Multiple working hypotheses in series (figure 1a) may, from the perspective of the observer, appear simultaneously true, yet may be separated in time by a sequence of state changes, with later actions and effects being dependent on former ones. In contrast, multiple working hypotheses in parallel (figure 1b) may, in practice, indeed be simultaneously true, and operate either independently or in interaction. This difference is important when considering how researchers might evaluate such hypotheses statistically, because multiple working hypotheses in series may be more readily distinguishable from each other as a result of their separation in time, and may thus be more easily approached by methods that test hypotheses one at a time. And although contemporary methods that explicitly accommodate the simultaneous comparison of hypotheses (e.g., model selection) may be applicable to both types of causation, they may be particularly well suited to scenarios in which multiple factors operate in parallel (figure 1b).

As a real-world ecological example of multiple factors working in parallel, Allan C. Fisher Jr. (1980) described a scene of rapid ecological change in Chesapeake Bay, the largest estuary system on the eastern coast of the United States. The bay serves as a hydrological mixing bowl, receiving fresh water from a number of tributaries and tidal salt water from the sea. Between 1968 and 1980, oyster sets in the lower James River were observed to decline from 2000 to 200 oysters per bushel, at a time when the rate of effluent in the tributaries had increased substantially. This effluent consisted of raw and chlorinated sewage, pesticides, herbicides and fertilizers from agriculture, heavy metals from industrial waste, and large volumes of sediment caused by erosion—excessive particulate matter that deprives oysters of oxygen for part of the year. Overly fresh water can also affect oysters, because they can only briefly tolerate saline solutions of less than five parts per thousand. In 1972, Hurricane Agnes wiped out over two million bushels and eliminated oysters entirely from some parts of the bay. There was also a devastating oyster disease, MSX (multinucleated sphere X [unknown]), which arrived in 1959 and, as the name implies, about which very little was known.

To answer the broad question of what caused the reduced rate of oyster set in the lower James River, there are a number of potentially interrelated factors to which we might attribute blame. It turns out that MSX is caused by a spore-forming protozoan (*Haplosporidium nelsoni*), which was found not to affect oyster larvae strongly because it cannot tolerate the relatively low salinity of the lower James. Further, it is possible to discount the hypothesis that Hurricane Agnes was responsible for the overall decline, because the oyster populations in the affected areas recovered after this event. But the other factors appear difficult to separate. Fisher (1980) concluded that "a combination of factors is putting the oyster larvae in great stress." And although only one factor (e.g., chlorine levels) may have actually been responsible, it is quite possible that several factors ("working hypotheses") acted simultaneously.

Chamberlin's method and Bradford-Hill's guidelines for causation. It is extremely difficult to prove causation in observational (and many experimental) studies, in part because there are usually many factors researchers cannot adequately control. As such, Austin Bradford-Hill (1966) instead developed a set of guidelines (later called "criteria" by others) for establishing causation in medicine and other fields (Phillips and Goodman 2004). These guidelines were used and accepted by the US Supreme Court in the case of *Daubert v. Merrill Dow Pharmaceuticals* (509 U.S. 579 [1993]), establishing legal precedent. On the basis of this decision, judges could deny the efficacy of defenses such as "There is no statistical evidence to prove that smoking causes lung cancer," which fail to acknowledge that investigators can use auxiliary information to infer whether causation is likely or to determine what is biologically plausible. Bradford-Hill did, however, overestimate people's ability to assess numer-
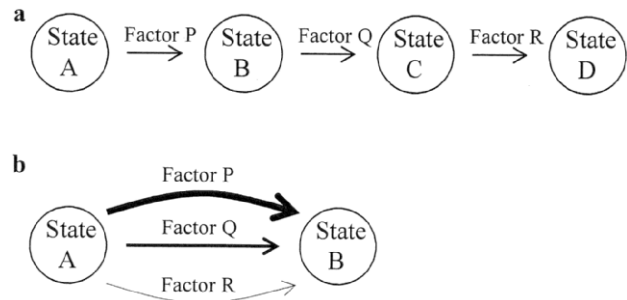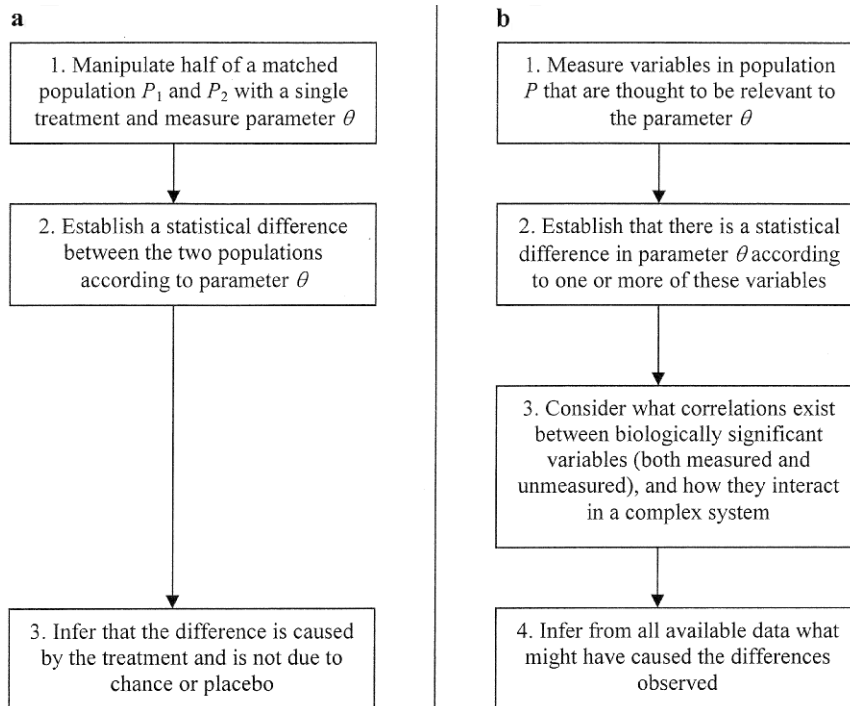

Figure 1. Comparison of two possible types of natural system where the method of multiple working hypotheses is applicable. Multiple factors can lead to a state transition both (a) in series (e.g., chains of extinction), where two or more factors occur sequentially, and (b) in parallel (e.g., ecosystem degradation), where the relative strength of simultaneous factors is indicated by the line thickness.

ical and probabilistic relationships; later work in the 1970s and 1980s demonstrated that laypeople have poor quantitative intuition (Phillips and Goodman 2004).

Chamberlin's description of the MMWH is concordant with Bradford-Hill's guidelines. "The effort," Chamberlin wrote, "is to bring up into view every rational explanation of new phenomena and to develop every tenable hypothesis respecting their cause and history" (Chamberlin 1890). Bradford-Hill (1966) suggested that a rational approach to establishing causation would include concepts such as "consistency, temporal sequence, coherence with biological background and previous knowledge, [and] biological plausibility." In this sense, Bradford-Hill and Chamberlin both describe a general method of inference as it applies to causation. Their concerns are specifically pertinent to step 3 in figure 2a ("Infer that the difference is caused by the treatment and is not due to chance or placebo") and steps 3 and 4 in figure 2b ("Consider what correlations exist between biologically significant variables...and how they interact in a complex system" and "Infer from all available data what might have caused the differences observed"). In a system in which a researcher can adequately control for extraneous factors and manipulate the factor under consideration, the pathway described in figure 2a, simple experimental design, is likely to be the more direct inferential route. Of course, many studies do not permit this luxury, and in these cases the process of inference must necessarily rely to a much greater extent on other kinds of knowledge (observational study; figure 2b).

"Strong inference" and the method of multiple working hypotheses

In the article "Strong Inference," John Platt (1964) clearly described the classic analytic method of science, first attributed to Bacon (1620). It requires a three-step process: (1) developing alternative hypotheses, (2) devising a set of crucial experiments to eliminate all but one of the hypotheses, and (3) performing the experiments. In its ideal representation,

a

1. Manipulate half of a matched population P_1 and P_2 with a single treatment and measure parameter θ

↓

2. Establish a statistical difference between the two populations according to parameter θ

↓

3. Infer that the difference is caused by the treatment and is not due to chance or placebo

b

1. Measure variables in population P that are thought to be relevant to the parameter θ

↓

2. Establish that there is a statistical difference in parameter θ according to one or more of these variables

↓

3. Consider what correlations exist between biologically significant variables (both measured and unmeasured), and how they interact in a complex system

↓

4. Infer from all available data what might have caused the differences observed

Figure 2. Two possible pathways to inference in ecological systems: (a) simple experimental design and (b) observational study.

the method is unidirectional, because progress is built on what has been tested, through the systematic growth and pruning of branches on a tree of scientific knowledge.

Platt (1964) argued, however, that the doctrine of disproof and falsifiability tends to force scientists to be either "soft-headed or disputatious." His implication was that there is a deficiency in the way scientists conduct their affairs when they proceed by negatives, a process that can lead to combative thought. Platt suggested that to be overly contentious in science can be counterproductive, because it makes some people shy away from using the benefits of the classic analytic method. Others are left vulnerable to the workings of the ego and to the risks of becoming emotionally attached to their ruling hypotheses, an outcome with its own inherent dangers to the spirit. Platt's solution was the application of Chamberlin's MMWH, and he called for the reprinting of Chamberlin's 1890 paper "where it could be required reading for every graduate student—and for every professor" (Platt 1964). However, although there are aspects of Chamberlin's method that are compatible with the classic analytic method as expressed by Platt (1964), there are also important differences that were not fully appreciated at the time.

Differences between Chamberlin's method and the classic analytic method. The key difference is that Chamberlin's MMWH recognizes explicitly the possibility that more than one hypothesis may be simultaneously true, while the classic analytic method, as espoused by Platt (1964), recognizes only that there is uncertainty as to which hypothesis repre-

sents truth. This does not imply that the use of the MMWH necessitates the introduction of complexity where a simple explanation will suffice. Rather, if more than one cause can be shown to exist, then the question becomes, What is their relative importance, and how do they interact? Further, under the MMWH these hypotheses are not framed as alternatives to be falsified in order to provide material for Bacon's method of inference. Indeed, Chamberlin's MMWH does not anticipate, and would not allow, crucial experiments to point in the direction of any single hypothesis. In this it distinguishes itself from the classic analytic method, which, in its ideal representation, requires (a) an exhaustive set of hypotheses and (b) a decisive falsification of all but one of these. In ecology, both conditions are difficult to meet in practice and, arguably, by its own construction, this method discourages synthesis and a consideration of multiple effects.

On the cultivation and invention of knowledge. Although it is not convenient here to be dogmatic about affiliations between positions in philosophy and positions in statistics, null-hypothesis testing is clearly based on Bacon's systematic method of inductive reasoning and decisive falsification of hypotheses (Platt 1964). Yet Bacon did not intend that the method of "the interpretation of nature" be the whole truth, but rather only one part of it:

Let there be therefore (and may it be for the benefit of both) two streams and two dispensations of knowledge, and in like manner two tribes or kindreds of students in philosophy—tribes not hostile or alien to each other, but bound together by mutual services; let there in short be one method for the cultivation, another for the invention, of knowledge. (Bacon 1620)

An initial interpretation might be that Bacon is describing the difference between science and nonscience. To the contrary, Bacon emphasizes that by the method of inductive reasoning, one may discover and demonstrate new knowledge in both the sciences and the arts (Platt 1964). This broader notion of science is accorded cultural universality by Colin Scott:

If one means by science a social activity that draws deductive inferences from first premises, that these inferences are deliberately and systematically verified in relation to experience, and that models of the world are reflexively adjusted to conform to observed regularities in the course of events, then, yes, Cree hunters practice

science—as surely as all human societies do.
(Scott 1996)

Scott refers not only to the testing of knowledge by way of experience but also to the adjustment of models of the world. This is a critical observation because it recognizes that there is much more to the process of science than is contained solely in Bacon's method for the invention of knowledge. Science includes the synthesis of different kinds of knowledge into consistent structures, and the use of imagination in developing explanations to account for our observations of the world. Some people call this critical or lateral thinking, and it is fundamental to all scientific disciplines. This is important because the greatest value of Chamberlin's MMWH (and also Bradford-Hill's guidelines for inferring causation) lies in the construction of hypotheses and the testing of complex systems in settings where explanations are not necessarily mutually exclusive.

Truth, null-hypothesis testing, and model selection. The notion of truth, in its various guises, is integral to the way different philosophers of science have approached the concept of hypothesis testing. However, pragmatic individuals (including most scientists) find it more convenient simply to get on with the job at hand, rather than to philosophize about such nebulous matters. So why consider it at all? We suggest two practical reasons: First, because good scientific method relies on the proper construction and testing of hypotheses, including the concept of falsification; second, because the way scientists justify their ability to make inferences from data is mediated by the branch of mathematics known as probability theory. It is here that notions of absolute and relative (or probabilistic) truth gain heightened importance in the way that we present data and derive our conclusions.

All methods of scientific inference place great importance on the proper construction of hypotheses in order to better approach the true state of nature. But in the classic analytic method, these hypotheses cannot, in fact, overlap in their design or content, because they need to be logically or statistically distinguishable from each other. Similarly, null-hypothesis testing requires that there be a defined condition that a hypothesis may or may not fulfill, and falsification consists of disproving hypotheses that are shown to be inconsistent with this existent truth (Anderson et al. 2000). Furthermore, these alternative hypotheses are not directly tested in any way—support is engendered by rejecting the null, and then inferring a plausible explanation. The Bayesian information criterion (BIC) is a dimension-consistent form of model selection intended to provide a measure of the weight of evidence favoring one model over another (the Bayes factor). The target of the BIC is the "true model," under the assumption that it is included in the model set under consideration. As the sample size becomes larger, the BIC approaches the estimation of the dimension of this true model with a probability of 1 (Burnham and Anderson 2004). In stark contrast, model selection based on information theory (usually applied using

Akaike's information criterion, or AIC) immediately states that *all* models are in fact false, because they represent incomplete approximations of a real but unreachable truth.

As an approximation of Kullback-Leibler information, AIC model selection weights models in accordance with their fit to the observed data, and represents the relative distances between conceptual reality and a set of approximating models (Burnham and Anderson 2002). Parsimony (essentially, Occam's razor) is reinforced as a result of the correction for asymptotic bias. In appraising simultaneously how different models fit with observed data based on "predictive likelihood" (i.e., out-of-sample prediction), a picture is developed that is less susceptible to the idea that a single model is in fact "true" (Burnham and Anderson 2004)—recall Chamberlin's description of the single working hypothesis. We argue that this position is more compatible with Chamberlin's position, and that the different philosophical basis of AIC model selection becomes a practical asset when assessing complex systems. The information-theoretic approach also drives the evolution of scientific hypotheses, because through repeated exposure to new data, hypotheses or models lacking any empirical support can be dropped, hypotheses remaining may be further refined, and new hypotheses will be derived and added to the working set.

This is not to suggest (as others have) that null-hypothesis testing methods should not be used. Despite the many articles written justly decrying the inappropriate use of *P* values (Carver 1978, Johnson 2002), there is nothing actually incorrect about using null-hypothesis testing methods to make inferences, and we do not propose to argue otherwise (see Stephens et al. [2005] for a recent discussion). It is simply that *P* values do not represent a proper "strength of evidence" (Lukacs et al. 2007). Yet there is a greater natural concordance (and potentially greater efficiency and economy of thought) between AIC model selection and the MMWH, and conversely between null-hypothesis testing, BIC model selection, and the classic analytic method. This may not, however, be a straightforward relationship; for instance, the MMWH in parallel (figure 1b) may be more suited to the use of AIC model selection than the more conceptual MMWH in series (figure 1a).

With the exponential rise in computing power, other alternative but numerically intensive paradigms, such as crossvalidation and full Bayesian inference (not BIC), are being used more frequently in ecology (Turchin 2003, Clark 2005). A Bayesian definition of probability includes the degree of belief in an event or model, an approach that enables greater flexibility when evaluating data from complex or incomplete data sets. With regard to Chamberlin's MMWH, Bayesian methods, like AIC model selection, have the distinct advantage of removing one's reliance on the literal falsification of competing hypotheses; they also allow for an explicit incorporation of uncertainty in the modeling process and in the accumulation of knowledge. It has been suggested that this shift toward alternative methods of inference has occurred primarily for pragmatic reasons (Stephens et al. 2005). We argue that a shift in philosophical position may be a natural

outcome of these recent developments. This has implications not only for the way that scientists consider the use of statistics when we make inferences about the world, but also for the way that science defines itself in relation to other kinds of knowledge.

Applications of Chamberlin's method and multimodel inference.

Model selection has been readily adopted in some areas of biology, especially wildlife management (e.g., capture–mark–recapture analysis to determine the interplay of intrinsic and environmental influences on survival and density), population ecology (e.g., establishing the relative plausibility of cyclic versus chaotic dynamics in long-term time series), and, increasingly, conservation biology (e.g., determining which anthropogenic factors best predict range declines in threatened species; for details on these and other biological examples, see Buckland et al. 1997, Hilborn and Mangel 1997, Morris and Doak 2002, Turchin 2003, Johnson and Omland 2004). A primary motivation for the application of model selection in these fields is the separation of biologically important signals from the myriad of "tapering effects" that characterize full truth but defy reductionism. An additional intellectual step, in particular harmony with sequential and simultaneous multiple working hypotheses (figure 1), is to invoke the concept of multimodel inference (MMI; Burnham and Anderson 2002). Rather than selecting a single "best" model, MMI involves making inferences on the basis of all models in an a priori candidate set, with the weighted contribution of each model (hypothesis) governed by its relative support from the data (estimated using scaled differences in AIC [AIC weights] or Bayesian posterior probabilities). MMI has three clear advantages: (1) It accounts explicitly for uncertainty in choosing the Kullback-Leibler best model (e.g., due to finite sample size), (2) it permits inference from different models that may be concurrently true (to lesser or greater extents), and (3) it allows researchers to estimate unconditional measures of precision and of the relative importance of variables.

A key example serves to illustrate these benefits. A longstanding debate has raged in population ecology over the appropriate means to detect intrinsic regulation (density dependence) in abundance time series (summarized in Brook and Bradshaw 2006). The classic approach has been to apply various significance tests, on a case-by-case basis, and evaluate whether there is sufficient evidence to reject a null model of density independence at $\alpha = 0.05$. But why, when there exists ample mechanistic evidence for the action of density feedbacks on survival and reproductive rates (Osenberg et al. 2002), should the null model be so favored? At the very least, the "competing" hypotheses of density independence versus density dependence should be evaluated on equal starting terms (a Bayesian might argue for strongly favoring density dependence a priori). Brook and Bradshaw (2006) used model selection to do this for time series covering 1198 species, and showed density dependence to be the better-supported hypothesis in 75 to 92 percent of cases, compared

with 33 to 50 percent for null-hypothesis testing. But perhaps more important, there were many instances in which there was reasonable strength of evidence for both hypotheses according to AIC, even for long-monitored populations with high statistical power. This supports a philosophical stance in which the question is not "Does density dependence occur in this population?" (in the broad sense of the classic analytic model), but rather "What is the relative importance of density-independent (extrinsic) and density-dependent (intrinsic) processes in driving the dynamics of this population?" This latter is a richer and more biologically meaningful line of inquiry, and more closely allied with the spirit of both MMI and Chamberlin's MMWH.

Conclusions

Western science and philosophy have a long tradition of thought concerned explicitly with the notions of observation, inference, truth, and prediction. Yet the statistical methods that we would recognize today are less than 100 years old. This raises a question: What were Hooke, Linnaeus, Cuvier, and Darwin doing before the development of the *P* value? Their substantial contributions to biology remind us that analytic thought is only one component of science. We posit that because null-hypothesis testing has been established as orthodox practice, the core meaning of Chamberlin's MMWH has been lost or altered. It now takes on even greater value than it had before, because it describes explicitly the processes of synthetic thought useful for approaching complexity, a timely concept for the application of 21st-century statistics.

Hilborn and Mangel (1997) liken the study of ecology to the investigations of a detective, whereby a coherent picture must be built up from an array of small and varied clues. It is no coincidence that Chamberlin's MMWH drew from a similar quarter—the field of geology—where it is difficult (and often inappropriate) to separate confounding variables using a series of dichotomous tests against a null. In fact there is a spectrum of possible approaches to analyses in natural science, from repeatable experimental designs with controlled treatments to diachronic observational studies in which the luxury of control is simply not tenable (Stephens et al. 2005). Chamberlin (1890) espouses a worldview that values lateral thinking and multiple possibilities. This can only be a positive development in a world where religious and political fundamentalism represents complexity in a black-and-white fashion. It seems appropriate that we revisit these ideas at such a time in history.

Acknowledgments

References cited

Anderson DR, Burnham KP. 2002. Avoiding pitfalls when using information–theoretic methods. Journal of Wildlife Management 66: 912–918.

Anderson DR, Burnham KP, Thompson WL. 2000. Null hypothesis testing: Problems, prevalence, and an alternative. Journal of Wildlife Management 64: 912–923.

Bacon F. 1620. The New Organon, or True Directions Concerning the Interpretation of Nature. (11 May 2007; *www.constitution.org/bacon/nov_org.htm*)

Bradford-Hill A. 1966. The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine 58: 295–300.

Brook BW, Bradshaw CJA. 2006. Strength of evidence for density dependence in abundance time series of 1198 species. Ecology 87: 1445–1451.

Buckland ST, Burnham KP, Augustin NH. 1997. Model selection: An integral part of inference. Biometrics 53: 603–618.

Burnham KP, Anderson DR. 2002. Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach. New York: Springer.

———. 2004. Understanding AIC and BIC in model selection. Sociological Methods and Research 33: 261–304.

Carver RP. 1978. The case against statistical significance testing. Harvard Educational Review 48: 378–399.

Chamberlin TC. 1890. The method of multiple working hypotheses. Science 15: 92–96 (reprinted in Science 148: 754–759 [1965]).

Clark JS. 2005. Why environmental scientists are becoming Bayesians. Ecology Letters 8: 2–14.

Fisher AC Jr. 1980. My Chesapeake—Queen of Bays. National Geographic 158: 428–467.

Hilborn R, Mangel M. 1997. The Ecological Detective: Confronting Models with Data. Princeton (NJ): Princeton University Press.

Johnson DH. 2002. The role of hypothesis testing in wildlife science. Journal of Wildlife Management 66: 272–276.

Johnson JB, Omland KS. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19: 101–108.

Lukacs PM, Thompson WL, Kendall WL, Gould WR, Doherty PF, Burnham KP, Anderson DR. 2007. Concerns regarding a call for pluralism of information theory and hypothesis testing. Journal of Applied Ecology 44: 456–460.

McBride GB, Loftis JC, Adkins NC. 1993. What do significance tests really tell us about the environment? Environmental Management 17: 423–432.

Morris WF, Doak DF. 2002. Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis. Sunderland (MA): Sinauer.

Osenberg CW, St. Mary CM, Schmitt RJ, Holbrook SJ, Chesson P, Byrne B. 2002. Rethinking ecological inference: Density dependence in reef fishes. Ecology Letters 5: 715–721.

Phillips CV, Goodman KJ. 2004. The missed lessons of Sir Austin Bradford Hill. Epidemiological Perspectives and Innovations 1: 3. (4 June 2007; *www.epi-perspectives.com/content/1/1/3*)

Platt JR. 1964. Strong inference. Science 146: 347–353.

Scott C. 1996. Science for the West, myth for the rest? Pages 69–86 in Nader L, ed. Naked Science: Anthropological Inquiry into Boundaries, Power, and Knowledge. New York: Routledge.

Stephens PA, Buskirk SW, Hayward GD, Martinez del Rio C. 2005. Information theory and hypothesis testing: A call for pluralism. Journal of Applied Ecology 42: 4–12.

Turchin P. 2003. Complex Population Dynamics: A Theoretical/Empirical Synthesis. Princeton (NJ): Princeton University Press.

Wade PR. 2000. Bayesian methods in conservation biology. Conservation Biology 14: 1308–1316.