frontiers in Ecology and Evolution

PERSPECTIVE
published: 09 August 2019
doi: 10.3389/fevo.2019.00306

Evidential Statistics in Model and Theory Development

Samuel M. Scheiner[1]* *and Robert D. Holt*[2]

[1] Division of Environmental Biology, National Science Foundation, Arlington, VA, United States, [2] Department of Biology, University of Florida, Gainesville, FL, United States

Evidential statistics is an important advance in model and theory testing, and scientific reasoning in general, combining and extending key insights from other philosophies of statistics. A key desiderata in evidential statistics is the rigorous and objective comparison of alternative models against data. Scientific theories help to define the range of models which are brought to bear in any such assessment, including both tried and trusted models and risky novel models; such theories emerge from a kind of evolutionary process of repeated model assessment, where model selection is akin to natural selection acting both on the standing crop of genetic variation, and on novel mutations. The careful use of evidential statistics could play an important and as yet to be fulfilled role in the future development of scientific theories. We illustrate these ideas using examples from ecology and evolutionary biology.

Keywords: abduction, deduction, evidential statistics, induction, model, theory

OPEN ACCESS

Edited by:
Mark Louis Taper,
Montana State University System,
United States

Reviewed by:
Brian Dennis,
University of Idaho, United States
Jeff E. Houlahan,
University of New
Brunswick Fredericton, Canada

Correspondence:
Samuel M. Scheiner
sscheine@nsf.gov

Specialty section:
This article was submitted to
Environmental Informatics,
a section of the journal
Frontiers in Ecology and Evolution

Received: *18 March 2019*
Accepted: *30 July 2019*
Published: *09 August 2019*

Citation:
Scheiner SM and Holt RD (2019)
Evidential Statistics in Model and
Theory Development.
Front. Ecol. Evol. 7:306.
doi: 10.3389/fevo.2019.00306

INTRODUCTION

Statistical inference aims at relating models to data and the empirical world, whether that model deals with an issue as simple as estimating the mean of a population or as complex as predicting millennial-scale changes in the global climate. There have been decades-long debates about the best way to make inferences (e.g., Neyman-Pearson error statistics vs. Bayesian approaches). This special feature highlights the approach called "evidential statistics," (Taper and Ponciano, 2016) which synthesizes prior approaches—error statistics, Bayesian statistics, information-based model selection, and likelihood approaches—to squarely focus on the comparative ability of alternative models or hypotheses for explaining an observed dataset. This approach to inference was sparked by Royall (1997) and Lele (2004), and the articles in this Special Issue highlight the rapid emergence and maturation of evidential statistics. We heartily concur with the value of such a synthesis of prior approaches, and the explicit emphasis on comparisons among alternative hypotheses or models as an essential component of scientific progress. Neither of us are card-carrying statisticians or philosophers of science; instead we are scientists interested in the conceptual basis of our discipline. Here we reflect on the need for intellectual flexibility by considering the role of statistical inference as a formal, mathematical procedure for refereeing the relationship between data, models, and theories, and place that in the context of the wider set of processes that scientists might use for theory development.

Scientists quest to obtain knowledge about the empirical world so as to understand its causal structure, and to use that causal structure for prediction as well as control and management. The inferential procedures employed to gain such knowledge should be "truth-tropic" (Lipton, 2004, p. 7). There are philosophers (e.g., Laudan, 1981) who reject the notion that science involves a kind of convergence toward an understanding of how nature works (conceived broadly), but we

feel that most working scientists assume (or at least hope) that they are engaged in a "truth-tracking" enterprise (Roush, 2007). While models are the direct connection between data and specific conclusions drawn from those data, those models are embedded within larger conceptual frameworks, typically called theories. One role of theory is to help guide the creative formulation of novel models for comparison against any set of data. For example, we might construct a family of ecological niche models (ENMs, Holt, 2009; Peterson et al., 2011) to explain why saguaro cacti (*Carnegiea gigantea*) are common in parts of the Sonoran Desert, yet absent elsewhere with seemingly comparable climates. Those models would be embedded within, and get their warrant from, broader theories of ecology and evolution (Scheiner and Willig, 2011b, Scheiner and Mindell, 2019). The models might draw upon diverse data and models such as the physiology of plants with Crassulacean Acid Metabolism as their mode of photosynthesis, the geographic history of North America, and the phylogeny of the Cactaceae. A criterion for selecting among alternative ENMs might be the minimization of errors in predicting known occurrences from available distributional and environmental data.

Statistics is essential for testing models in the broad sense, examining their relationship with the empirical world, efforts that in turn contribute to the goal of crafting and testing more general theories. Building and testing theories relies on a variety of approaches, only some of which make explicit use of statistical inference. Evidential statistics aims at providing a systematic approach for assessing the relative informativeness of models, which depends upon available data and protocols—distinct from the personal beliefs embedded within Bayesian statistics—via objective metrics of evidence that ideally lead toward closer approximations of the "truth" as models continue to be refined and compared (Dennis et al., 2019). Theories are distillations of conclusions (Tukey, 1960) achieved collectively by scientists, carrying out such protocols repeatedly and objectively. Kuhn (1977, pp. 321–322) notes that the development of scientific theories must juggle qualities which at times may be contradictory, such as accuracy, consistency, simplicity, fruitfulness, and scope, to which Houlahan et al. (2017) add as an essential desideratum the successful prediction of novel states of the world.

Like any evolutionary process, theory development depends upon the availability of an array of alternative models for comparison, using both a standing crop of existing models that have proven useful in other contexts, and novel conceptual mutations. Evidential procedures are akin to natural selection culling genetic variants, favoring the fittest in the population at hand in a given environment. For example, in our saguaro cactus model, general climatic variables such as average rainfall or seasonal patterns in precipitation are doubtless important and would discriminate among many models, but a key idiosyncratic factor operating at the northern range limits appears to be the number of consecutive hours below freezing (MacArthur, 1972, p. 127), which can be strongly influenced by local topography. The fittest of the competing models would surely need to include this key observation.

This evolutionary perspective on theory development stems back to Popper (1972, p. 261) who states, "[T]he growth of our knowledge is the result of a process closely resembling what Darwin called 'natural selection,' that is, *the natural selection of hypotheses*." In a sense, likelihood and related quantitative approaches provide fitness metrics for selecting some hypotheses over others based on evidence. Just as natural selection does not comprise all of evolution, knowledge development leading up to a general theory is more than just the accumulation of episodes of such evidence-based selection. Other processes, such as intellectual coherence, the generation of novel ideas, and the infusion of ideas across disciplinary boundaries, play roles comparable to mutation, gene flow, and recombination. A particular challenge is to articulate how the scientific community builds larger arenas of knowledge—theories—from more specific models grounded in evidence. Popper (1972, p. 262-3) suggests a kind of inverse evolutionary tree of knowledge emerging over time: "[T]he tree of knowledge [springs up] from countless roots which grow up into the air rather than down, and which ultimately, high up, tend to unite into one common stem." We now turn our attention to the relationship between models and theories, broadly conceived.

FROM MODELS TO THEORIES AND BACK AGAIN

Our approach to models and theories can be considered part of the Pragmatic View of the structure of scientific theory (Winther, 2012, 2015). The Pragmatic View combines formal components of mathematic axioms and associated models with less formal, non-mathematical components including concepts, metaphor, narrative, and analogy. The result is a pluralistic and pragmatic structure for scientific theory in which theory content is organized according to the research questions being asked (Love, 2010). Vandermeer (2018), in an encomium to Richard Levins, cogently remarks on why in biology, theory is not just a compilation of models: "Populations of organisms only approximately follow precise equations and theories about them thus cannot rely exclusively on models… [and] [m]athematical forms of models are tools, as Levins repeatedly expressed, 'to educate the intuition.'"

Scheiner and Willig (2008) proposed a hierarchical framework for organizing theories consisting of general theories, more narrow constitutive theories, and even more specific models. The three types of theories have different functions. General theories provide the conceptual framework within which theories and models are built and tested. They consist of a set of general principles—confirmed generalizations—that provide background assumptions. These principles may appear trivial, but that is only because they have been so thoroughly tested that they have become embedded in our background knowledge. Yet, they are often ignored when building models. For example, one of the general principles of the theory of ecology is that "Variation in the characteristics of organisms results in heterogeneity of ecological patterns and processes" (Scheiner and Willig, 2011a). It is a reminder that even though very many ecological models

assume that all individuals within a species are identical, we know that this is an approximation. While violations of this assumption may not substantially change model predictions in some situations, in other cases relaxing this assumption even by a small amount can lead to marked changes (e.g., Kendall and Fox, 2003). The constitutive theories and models are not derived formally from general theories. Rather, general theories provide the background knowledge and general conceptual framework within which more specific theories and models are built. For more on this conceptualization of a theory hierarchy (the inverse knowledge tree of Popper, 1972), see Scheiner (2010) and Mindell and Scheiner (2019).

Constitutive theories are the workhorses in this framework and what most individuals would think of when asked to name or describe a theory. Their role is to organize models into larger entities. They consist of a set of propositions, which might arise inductively from a set of models (e.g., a constitutive theory of diversity gradients, Scheiner and Willig, 2005). Alternatively, the propositions might be conceived first and then used to guide model development (e.g., the theory of natural selection, Frank and Fox, 2019). For example, enemy-victim theory (Holt, 2011) includes, among others, three propositions: (1) The increased consumption generated by increased victim abundance in turn fuels an increase in the per capita growth rate (fitness) of the natural enemy population. (2) An increase in the victim population increases the rate of consumption by each individual natural enemy. (3) Consumption by the natural enemy implies mortality in the victim. Making simplifying assumptions about the functional forms for each of these (which in turn reflect models and theories about the component processes), along with ancillary assumptions (e.g., no direct density dependence), these propositions can be formalized as the classical Lotka-Volterra predator-prey model:

$$\frac{dP}{dt} = P[baN - m]$$
$$\frac{dN}{dt} = N[r - aP]$$

where P and N are the densities of predators and prey, respectively, the predator birth rate is given by baN, where a is the attack rate and b is the rate that prey biomass is converted into offspring, m is the predator death rate, and r is the prey birth rate. This model is just one particular instantiation of those propositions; many other versions are possible. These models then serve to link theories to data, which is where evidential statistics comes into play.

The framework is multilayered, and both general and constitutive theories can be nested and overlapping. For example, a model of the evolution of plasticity of *Drosophila melanogaster* body size in response to temperature is embedded within a constitutive theory of the evolution of phenotypic plasticity that draws upon the constitutive theory of evolution by natural selection, both in turn embedded within the theory of evolution (Scheiner, 2019), while also drawing upon constitutive theories within the theory of organisms (Zamer and Scheiner, 2014). Some of these constitutive theories include formalized mathematical models, but others do not.

Models can be both qualitative and quantitative in describing or predicting nature. In ecology and evolution we tend to think of dynamical mathematical models, systems of equations or computer rules linked by logical operators corresponding to assumptions about mechanisms at and across different levels of biological organization. A computer simulation, such as an individual-based model of population dynamics, might be an example. Models can also be qualitative; Charles Darwin's theory of evolution was almost entirely verbal and qualitative. There is a single, iconic tree-like figure in *On the Origin of Species* which displays the grand, overarching vision of a shared origin for all life in an instantly transparent manner—an elegant example of a graphical, non-mathematical model.

From models we deductively derive hypotheses that in turn make predictions. These predictions are often derived from a mathematical model, which are based on some expected distribution of parameter values (see other articles in this special feature). Those distributions are then compared to data (broadly defined). Whereas the model is general in the sense that it applies across a domain of interest, a hypothesis becomes a prediction when applied to a specific, empirical instance. That application, the collision of models and data, is where evidential statistics steps in.

THE RELATIONSHIP OF EVIDENTIAL STATISTICS TO MODELS AND CONSTITUTIVE THEORIES

Statistical methods shed light on the possible relative verisimilitude or falsity of a hypothesis, compared to coherently-specified alternative hypotheses. That hypothesis might be that a model parameter has a very specific value (e.g., in plant populations the relationship between the average mass per individual and the density of survivors should have a exponent of −3/2, Yoda et al., 1963), or it could be more general (e.g., the relationship between productivity and diversity is hump-shaped, VanderMeulen et al., 2001), or it could be qualitative (e.g., the mating system in this particular plant population will be gynodioecy). By inference, if the hypothesis is false then the model is inadequate in the sense that compared to some alternative model, the model in question does not correspond to the empirical world. The history of science is littered with failed models and hypotheses (e.g., phlogiston, the ether, epicycles, barnacles as larval stages of barnacle geese), and many scientific advances prove to be way stations toward a deeper understanding of the world (e.g., Newton's gravitational theory). But statistics does not have the same role (at least not so obviously) when it comes to constitutive or general theories.

Those theories are systems that organize models, data, concepts, and so forth [Box 3.2 in Pickett et al. (2007) describes the components of theories]. Considered as an organizational system, constitutive and general theories are never true or false. Rather, they are useful, not useful, or poorly structured, that is, conceptually fruitful or not. That is not to say that

general theories (e.g., the theory of evolution) are not true; rather that the strength of the theory lies in the overall validity of its components, rather than a single assessment of the entire theory.

Within constitutive theories are families of models, and decisions need to be made as to which models to include or exclude. Sometimes that decision-making process is how well one model mirrors the empirical world relative to another model. Evidence based on the relationship of a hypothesis with data and the empirical world leads to inferences about the relative truth or falsity of the hypotheses generated by each model, a decision-making process mediated by statistics. But these decisions are only part of what goes into conclusions about the utility of a constitutive or general theory. A principle in a general theory (e.g., "The ecological properties of species are the result of evolution" from the theory of ecology, Scheiner and Willig, 2011a) comes from the accumulation of a multitude of individual observations and models. An individual model can be discarded without negating the more general theory. We might decide that a natural selection model of the frequency of third position codons in DNA is inapplicable, because third position codons evolve by drift (Kimura, 1968). That conclusion would not affect the status of the theory of evolution by natural selection.

Evaluating models, such as the predator-prey model given above, involves more than just comparing predictions with data. That model famously predicts predator-prey cycles, looking in some respects like real-world cycles (such as the lynx-snowshoe cycle of Canada). May (1973) pointed out, however, that these models are neutrally stable, and so are highly unlikely to describe real cycles that are persistent. Indeed, the model is structurally unstable, in that small deviations in model assumptions lead either to oscillations that blow up, or to a stable equilibrium. Structural stability should be a desideratum in all our model and theory construction. Yet, real organisms and communities are unlikely to exactly match any set of equations we are likely to concoct.

Models may be false, while still playing a vital role in the conceptual framework of ecological theory. We contrast structural stability (the robustness of model conclusions to small deviations in model assumptions) with the stability of model structure. For the predator-prey model, the essential structure of the model itself (a +/– interaction between two antagonists, a natural enemy and its victim) is applicable across many empirical systems (e.g., predator-prey, host-pathogen, and plant-herbivore). The Lotka-Volterra predator-prey model demonstrates that there is a tendency to oscillate inherent in such antagonistic interactions. This qualitative conclusion is robust across many variants of this basic model, although the details may differ (e.g., the oscillations may manifest as transients following a perturbation, rather than as permanent cycles). Because the Lotka-Volterra model makes such robust, qualitative predictions, it continues to play an important role in the conceptual framework of theoretical ecology, even though it is known to be literally false for all empirical predator-prey systems. The same can be said of the model of exponential growth, $dN/dt = rN$, where N is population size and r is the intrinsic rate of increase. It has been argued that the principle of exponential growth is one of the conceptual foundations of ecology (Pásztor et al., 2016), and Ginzburg and Colyvan (2004) state that "the whole body of the spectacularly successful evolutionary theory has Malthusian growth in its foundation." Yet essentially no populations, when examined closely, match this model—there are always age and stage structure effects, demographic and environmental stochasticity, genetic variation, spatial dynamics, and density-dependent feedbacks, at play. This sweeping generalization, however, does not vitiate the conceptual role of exponential growth as foundational in our discipline. In like manner Queller (2017), commenting on Ronald Fisher's fundamental theorem of evolution, notes that it leaves out many important drivers of evolutionary change, but nonetheless "demonstrate[s] the general value of simplifying and sacrificing a bit of accuracy in order to capture and highlight fundamental issues in a simple and elegant way." This highlighting is an essential role of theory—enhancing understanding.

If a theory is relatively narrow, encompassing just one or a few specific models, and all of those models fail, we would then discard the theory as not useful. For example, Arditi and Ginzburg (2012) argued that we should discard any theory of predation in which the rate that predators attack prey depends only on prey density but not predator density, as in the above Lotka-Volterra model, which illustrates what is called "prey-dependence." They compiled case studies that included formal estimates of a key parameter (m) measuring the strength of predator interference on foraging rates. While they did not do a formal meta-analysis of those estimates, if they had, statistical inference would likely have supported the conclusion that this effect of predator density needs to be incorporated into any predatory-prey model (but see Abrams, 2015). There is a large body of food web and network theory that simply assumes prey-dependence in trophic linkages (i.e., ignores predator density). It is not yet known if altering this assumption would merely tweak the rich body of conclusions drawn from this theory, or instead if the change would have revolutionary effects on ecological understanding.

The use of statistics to assess hypotheses and models involves both deductive and inductive reasoning. We deduce hypotheses/predictions from a model. If a prediction proves false, one or more aspects of the model may be concluded to be false, which is the basis of Popper's (1959) falsifiability criterion for scientific theories. We also use statistics as a form of inductive reasoning. With induction, we infer a general conclusion from particular instances. When we estimate a population parameter from an observed set of data (e.g., the mean weight of a population of *Drosophila melanogaster*), we are performing induction. A constitutive or general theory includes a set of confirmed generalizations—condensations and abstractions, ultimately, from a body of facts—that may include parameter estimates (e.g., the base-pair mutation rate), used in particular model comparisons. Evidential statistics (Taper and Ponciano, 2016) is based upon rigorous comparisons of the likelihood (broadly conceived) of two or more alternative

models. But it does not specify where the set of alternative models come from in the first place. This is where constitutive and general theories come into play—representing a kind of closet collective Bayesianism, where the cumulative wisdom of scientists over time help define the range of models that are likely to be assessed against any given dataset (Longino, 2002), as well as providing a structure for the creation of novel models.

A third, less familiar, type of reasoning is abduction. The term was coined by Charles Peirce (Douven, 2017), who used it initially to encompass hypothesis generation, but later in a manner related to the idea of "inference to the best explanation." The basic notion is that one compares alternative models and accepts the one that best explains the evidence. What counts as "best" could be its likelihood (in the sense used in evidential statistics as articulated by the other papers in this special feature), but also can involve desiderata such as simplicity, unification across studies, structural stability, and so forth (Lipton, 2004). Many of these ideas about how one can build up from models to more general theories can be traced to Whewell's (1858) three criteria for theory confirmation: prediction, consilience (explaining phenomena of a different kind than those used to formulate the theory), and coherence (the simplification or unification of different phenomena without the need for *ad hoc* modification of the theory) (Forster and Wolfe, 1999; Snyder, 2019). Norton (in prep, https://www.pitt.edu/~jdnorton/papers/material_theory/9.%20Best%20Explanation%20Examples.pdf) argues that Darwin's entire theory (as expressed in *On the Origin of Species*) involves an extended inference to the best explanation, all without explicit statistical inference. To our knowledge, no philosopher of science has yet brought together the notion of inference to the best explanation, and the complementary but distinct concepts of confirmation and evidence articulated by Bandyopadhya et al. (2016). Mark Taper (pers. comm.) notes that one virtue of evidential statistics is that one keeps track not just of the "best" model, but other models that might prove useful in future investigations. Evidential statistics provides a clear path for comparing models against particular datasets; what is now needed is an articulation of higher-order protocols for assessing constitutive and general theories. Such protocols are presumably at play when a community of scientists converge on particular ways of understanding the world. The bridge from models to more general theories may be more loosely constructed in biology than in, say, quantum physics. As Vandermeer (2018, p. 4) cogently notes, "[In population biology] any model is only approximate with respect to the theory it intends to represent, and any theory is bolstered by its conformation, even if approximate, to multiple models."

The development of constitutive and general theories cannot be entirely shoe-horned into formal statistical inference, including evidential statistics, vital though that is for sifting hypotheses and models. Statistical inference alone is insufficient when dealing with the sculpting over time of scientific understanding, involving the concerted efforts of many scientific minds who collectively craft complex models or theories (Longino, 2002). The total weight of the evidence that bears on theory development includes not just the quantification of specific estimated parameters, or alternative functional forms of models, but also reflects our confidence in the logical structure and explanatory scope of the models that are derived from a constitutive theory, and whether the domain of that theory encompasses the specific instances under consideration. In some sense, constitutive and general theories rely upon a higher order of evidential support and logical considerations that may lie outside the specific scope of any given dataset. For example, when examining a particular trait, such as emergence of blindness in a cave fish in Kentucky, should our models invoke only natural selection, or also the accumulation of deleterious mutations and genetic drift? The answer to this question would likely depend on what has been learned about other cave fish worldwide. Taper and Ponciano (2016) use Gause's (1934) famed protozoan experiments to compare the relative evidentiary power of a suite of population dynamic models, such as the Ricker, Beverton-Holt, and Gompertz equations. Choosing this suite of models for comparison, and excluding others, implicitly involves a priori beliefs about the relevant drivers of population dynamics, presumably drawing on correspondences between this concrete empirical system and a wide array of somehow comparable systems, as well as more specific assumptions, such as: there is no spontaneous generation, the populations are closed to immigration and emigration so that local births and deaths entirely drive dynamics (this is ensured by the experimental setup), there are no time-lags in density dependence (which might occur with the buildup of toxins or waste products, or subtle stage-structure effects), and there are no hidden players such as viruses. These background assumptions help define the range of models to be compared explicitly, using the metrics of evidentiary statistics.

What is the role of evidential statistics in determining the relationship between models and theories where the latter are qualitative, rather than quantitative? For example, our explanation about the range of saguaro cacti includes information about the geographic history of the North and South American continents. We have models of the movements of the continents over geological history, but those models are not mathematical equations. Rather, we have inferred that history from a range of observations, only some of which include quantitative models. In modern systematics, a phylogeny is a quantitative model of a set of relationships among species (or higher taxa) in a clade. When multiple phylogenies are overlain on a map, the subsequent qualitative biogeographic patterns can be used to make inferences about the geological history of that region. It is possible to devise a formal inference process for making decisions about that history, but a formal process is not always necessary. Wegener's (1966) theory of continental drift was based, in part, on observing close phylogenetic relationships between South American and African species, as well as the fit of the shapes of the continents themselves. This process of bringing together models from multiple domains that all point to the same explanation is an illustration of the concept of "consilience" first championed by Whewell (1840). Ferguson et al. (2012) provide an example of how to devise statistical inference procedures when both predictions and data are qualitative. It strikes us that this may be one arena ripe for further analysis and formalization.

WHERE WILL EVIDENTIAL STATISTICS GO, AND HOW BEST CAN IT BE USED TO INFORM AND REFINE CONSTITUTIVE AND GENERAL THEORIES?

Evidential statistics is still a relatively new approach to linking data, models, and constitutive theories, but it promises to provide a clearer and more coherent way to assess the relative match of models to data, compared to competitors such as Neyman-Pearson testing or Bayesian analysis. Does the use of evidential statistics change if the purpose of a model is for understanding (e.g., why saguaro are confined to the Sonoran Desert) vs. prediction (e.g., what is the most likely global mean temperature in the year 2100)? Does this use change if the model is mechanistic vs. phenomenological? Are different evidence functions better suited for prediction vs. explanation? If one carries out multiple studies, each of which uses evidence functions, how can these best be brought together to examine broad-scale patterns across many systems? Maybe there is a straightforward, evidentiary-statistics version of meta-analysis (for a start, see Goodman, 1989). We use statistical inference to find the model that best fits the data. But the better fitting model may be "less true," in the sense of providing less understanding. A more "accurate" model can be the result of overfitting, especially if the model is phenomenological. Some types of statistical inference (e.g., the use of information criteria like AIC and BIC) try to correct for the inclusion of unnecessary parameters, but we also rely on logical reasoning and prior information to decide which parameters and functional forms are even appropriate to include, a process that is outside of statistical inference itself. For instance, a mathematical model must have units on each side of the equal sign that match; if not the model is, at best, nonsense. A number of evidence functions have been proposed in the literature, and presumably the class of such functions will grow with time. Are the criteria used to assess those functions part of evidentiary statistics, or in some sense outside of it?

If the goal is understanding, a very simple model may be appropriate. For example, we might ask whether saguaro abundance within its occupied range is controlled by intraspecific competition only, or also by interspecific competition with ferrocactus. We could build a very simple model of logistic growth without and with competition and use inferential statistics to ask which model is more consistent with observed densities across space and/or time. The model is not likely to be useful for making an accurate prediction of densities, but may nonetheless help uncover the presence of a particular ecological mechanism (e.g., competition). Simple models can illuminate essential elements of a system, even if statistical inference indicates that the model is very far from an accurate depiction of the empirical system. Depending on our goal, the most useful model could either be very simple (to highlight a single, essential feature) or very complex (to be as accurate as possible). In this case, our goal is not theory testing. Rather, the goal is to use an established theory to build a model for a specific instance so as to enhance understanding.

Prediction is important and indeed vital in the progress of science (Houlahan et al., 2017), but it does not outweigh other considerations in theory evaluation. After all, geocentric Ptolemaic astronomy did a fine job of predicting the movement of the planets for over 1,500 years, at the expense of more and more model complexity. Its supplanting by a gravity-driven, heliocentric theory, was driven, in part, by the latter model being both mechanistic and much simpler. The excellence of Ptolemaic astronomy as a predictive tool is not a very strong argument for hanging on to it as science moves forward. Newton's remarkable accomplishment in his *Principia Mathematica* was to explain an array of already known facts—Kepler's laws, tidal rhythms, the precession of the equinoxes—using just his three laws of motion plus the inverse-square law of gravitation. Novel predictions eventually emerged (e.g., the existence of Neptune), but such predictions were not required for the scientific community by-and-large to become enthusiastic champions of Newtonian mechanics. The super-computers of the future are likely to use vast neural networks, evolving arrays of code-based algorithms, and constant training with the flood of informatics they are constantly fed from arrays of sensors and surveys, and the like, to provide wonderful predictions of climate change and the weather, but this will not substitute for causal, theoretical understanding, often relying at its core on models that are not literally true.

WHEN STATISTICS ARE NOT NECESSARY

Sometimes statistical inference is not necessary for testing a theory, for example when a model is being used to explore if something is possible or not. The data are simply that some object or phenomenon exists or does not exist. The model either matches the data or it does not; no statistical inference is needed. For example, contra the "central dogma" we might have a theory that acquired characteristics can be inherited. For over a century, all of the data said that this theory was false. Then retroviruses were discovered showing that information can flow from RNA acquired from the environment back to DNA. For at least this narrow domain, the theory of the inheritance of acquired characteristics has been shown to be true. One might be able to shoehorn such examples into evidential statistics, but it is not clear that is necessary to understand the logic of scientific discovery in cases of this sort.

Even with a question that is less clear cut than simply "Does it exist?" statistical inference may be unnecessary. Statistical inference is about finding the informative signal within noisy data. For highly controlled experiments, the noise might be so small that the signal is immediately obvious. We know physiologists who say that if you need to use statistics, you really should refine your experimental methodology. Statisticians sometimes refer to this as the interocular trauma test, as in "it hits you between the eyes." Mark Taper (pers. comm.) ripostes "[Y]ou are still comparing the fit of data to models – it is just that the integration can be done by eye." Our evolutionary history has presumably fit us to be pretty good seat-of-the-pants statisticians, in that our past inferences have helped our ancestors survive and reproduce. But this decision process is not the same as

the formal mathematics of statistical inference represented by evidential statistics.

CONCLUSION

Evidential statistics is an important advance in model and theory testing, and scientific reasoning in general, combining and extending key insights from other philosophies of statistics. We applaud the editors and authors of this special issue for crystallizing many important exciting themes swirling around the topic of evidential statistics. A scientist should use whichever tool is apt for the particular question at hand. Statistical inference itself is just one class of tools used in scientific inquiry that depends on quantitative data and mathematical reasoning. Other types of data and reasoning are sometimes more appropriate for a given question, such as qualitative data, and narrative or logical reasoning. We urge scientists to use as wide a range of tools as possible in the service of our quest to understand, predict, and manage our ever-fascinating, complex world.

AUTHOR CONTRIBUTIONS

The authors equally conceived of the content and wrote the paper.

FUNDING

RH was supported by the University of Florida Foundation. This manuscript is based on work done by SS while serving at the U.S. National Science Foundation. The views expressed in this paper do not necessarily reflect those of the National Science Foundation or the United States Government.

ACKNOWLEDGMENTS

We thank Mark Taper and Jeff Houlahan for their extensive and perceptive comments on earlier drafts that greatly helped to strengthen our presentation.

REFERENCES

Abrams, P. A. (2015). Why ratio dependence is (still) a bad model of predation. *Biol. Rev.* 90, 794–814. doi: 10.1111/brv.12134

Arditi, R., and Ginzburg, L. R. (2012). *How Species Interact*. New York, NY: Oxford University Press.

Bandyopadhya, P. S., Brittan G. Jr., and Taper, M. L. (2016). *Belief, Evidence, and Uncertainty: Problems of Epistemic Inference*. Zurich: Springer International Publishing AG.

Dennis, B., Ponciano, J. M., Taper, M. L., and, Lele, S. R. (2019). Errors in statistical inference under model misspecification: evidence, hypothesis testing, and AIC. *Front. Ecol. Environ.* 7.

Douven, I. (2017). "Abduction," in *The Stanford Encyclopedia of Philosophy*, ed E. N. Zalta (Metaphysics Research Lab, Stanford University). Available online at: https://plato.stanford.edu/archives/sum2017/entries/abduction/

Ferguson, J. M., Taper, M. L., Guy, C. S., and Syslo, J. M. (2012). Mechanisms of coexistence between native bull trout (*Salvelinus confluentus*) and non-native lake trout (*Salvelinus namaycush*): inferences from pattern-oriented modeling. *Can. J. Fish. Aquatic Sci.* 69, 755–769. doi: 10.1139/f2011-177

Forster, M. R., and Wolfe, A. B. (1999). *Conceptual Innovation and the Relational Nature of Evidence: The Whewell-Mill Debate*. Available online at: http://philosophy.wisc.edu/forster/papers/Whewell1.pdf

Frank, S. A., and Fox, G. A. (2019). "The inductive theory of natural selection," in *The Theory of Evolution*, eds S. M. Scheiner and D. P. Mindell (Chicago: University of Chicago Press), 171–193.

Gause, G. F. (1934). *The Struggle for Existence*. Baltimore, MD: The Williams & Wilkins Co.

Ginzburg, L., and Colyvan, M. (2004). *Ecological Orbits: How Planets Move and Populations Grow*. Oxford: Oxford University Press.

Goodman, S. N. (1989). Meta-analysis and evidence. *Control. Clin. Trials* 10, 188–204. doi: 10.1016/0197-2456(89)90030-5

Holt, R. D. (2009). Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19659–19665. doi: 10.1073/pnas.0905137106

Holt, R. D. (2011). "Natural enemy-victim interactions: do we have a unified theory, yet?" in *The Theory of Ecology*, eds S. M. Scheiner and M. R. Willig (Chicago: University of Chicago Press), 125–161.

Houlahan, J. E., McKinney, S. T., Anderson, T. M., and McGill, B. J. (2017). The priority of prediction in ecological understanding. *Oikos* 126, 1–7. doi: 10.1111/oik.03726

Kendall, B. E., and Fox, G. A. (2003). Unstructured individual variation and demographic stochasticity. *Conserv. Biol.* 17, 1170–1172. doi: 10.1046/j.1523-1739.2003.02411.x

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* 217, 624–626. doi: 10.1038/217624a0

Kuhn, T. S. (1977). *The Essential Tension*. Chicago: University of Chicago Press.

Laudan, L. (1981). A confutation of convergent realism. *Philos. Sci.* 48, 19–49. doi: 10.1086/288975

Lele, S. R. (2004). "Evidence functions and the optimality of the law of likelihood," in *The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations*, eds M. L. Taper and S. R. Lele (Chicago: University of Chicago Press), 191–216.

Lipton, P. (2004). *Inference to the Best Explanation, 2nd Edn*. London: Routledge.

Longino, H. E. (2002). *The Fate of Knowledge*. Princeton, NJ: Princeton University Press.

Love, A. C. (2010). "Rethinking the structure of evolutionary theory for an extended synthesis," in *Evolution - The Extended Synthesis,* eds M. Pigliucci and G. B. Müller (Cambridge, MA: MIT Press), 403–441.

MacArthur, R. H. (1972). *Geographical Ecology*. Princeton, NJ: Princeton University Press.

May, R. M. (1973). *Stability and Complexity in Model Ecosystems*. Princeton, NJ: Princeton University Press.

Mindell, D. P., and Scheiner, S. M. (2019). "The theory of evolution," in *The Theory of Evolution,* eds S. M. Scheiner and D. P. Mindell (Chicago: University of Chicago Press), 1–22.

Pásztor, L., Bottak-Dukát, Z., Magyar, G., Czárán, T., and Meszena, G. (2016). *Theory-Based Ecology: A Darwinian Approach*. Oxford: Oxford University Press.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martinez-Meyer, E., Nakamura, M., et al. (2011). *Ecological Niches and Geographical Distributions*. Princeton, NJ: Princeton University Press.

Pickett, S. T. A., Kolasa, J., and Jones, C. G. (2007). *Ecological Understanding: The Nature of Theory and the Theory of Nature, 2nd Edn*. New York, NY: Elsevier.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. London: Hutchinson & Co.

Popper, K. R. (1972). "Evolution and the tree of knowledge," in *Objective Knowledge*, ed K. R. Popper (Oxford: The Clarendon Press), 133–143.

Queller, D. C. (2017). Fundamental theorems of evolution. *Am. Natural.* 189, 345–353. doi: 10.1086/690937

Roush, S. (2007). *Tracking Truth: Knowledge, Evidence, and Science*. Oxford: Oxford University Press.

Royall, R. M. (1997). *Statistical Evidence: A Likelihood Paradigm*. New York, NY: International Thompson Publishing.

Scheiner, S. M. (2010). Toward a conceptual framework for biology. *Q. Rev. Biol.* 85, 293–318. doi: 10.1086/655117

Scheiner, S. M. (2019). "The theory of the evolution of plasticity," in *The Theory of Evolution,* eds S. M. Scheiner and D. P. Mindell (Chicago: University of Chicago Press), 254–272.

Scheiner, S. M., and Mindell, D. P. (2019). *The Theory of Evolution*. Chicago: University of Chicago Press.

Scheiner, S. M., and Willig, M. R. (2005). Developing unified theories in ecology as exemplified with diversity gradients. *Am. Natural.* 166, 458–469. doi: 10.1086/444402

Scheiner, S. M., and Willig, M. R. (2008). A general theory of ecology. *Theor. Ecol.* 1, 21–28. doi: 10.1007/s12080-007-0002-0

Scheiner, S. M., and Willig, M. R. (2011a). "A general theory of ecology," in *The Theory of Ecology,* eds S. M. Scheiner and M. R. Willig (Chicago: University of Chicago Press), 3–19.

Scheiner, S. M., and Willig, M. R. (2011b). *The Theory of Ecology*. Chicago: University of Chicago Press.

Snyder, L. J. (2019). "William Whewell," in *Stanford Encyclopedia of Philosophy*, ed E. N. Zalta. Available online at: http://plato.stanford.edu/archives/spr2019/entries/whewell/

Taper, M. L., and Ponciano, J. M. (2016). Evidential statistics as a statistical modern synthesis to support 21st century science. *Popul. Ecol.* 58, 9–29. doi: 10.1007/s10144-015-0533-y

Tukey, J. W. (1960). Conclusions vs. decisions. *Technometrics* 2, 423–433. doi: 10.1080/00401706.1960.10489909

Vandermeer, J. (2018). "Objects of intellectual interest have real life impacts: the ecology (and more) of Richard Levins," in *The Truth is the Whole: Essays in Honor of Richard Levins,* eds T. Awebuch, M. S. Clark, and P. J. Taylor (Arlington, MA: The Pumping Station), 1–9.

VanderMeulen, M. A., Hudson, A. J., and Scheiner, S. M. (2001). Three evolutionary hypotheses for the hump-shaped productivity-diversity curve. *Evol. Ecol. Res.* 3, 379–392.

Wegener, A. (1966). *The Origin of Continents and Oceans, 4th Edn*. New York, NY: Dover Publications.

Whewell, W. (1840). *The Philosophy of the Inductive Sciences, Founded Upon Their Histories*. London: John W. Parker.

Whewell, W. (1858). *Novum Organon Renovatum*. London: J. W. Parker and Son.

Winther, R. (2012). Mathematical modeling in biology: philosophy and pragmatics. *Front. Plant Sci.* 3:102. doi: 10.3389/fpls.2012.00102

Winther, R. G. (2015). "The structure of scientific theories," in *Encyclopedia of Philosophy,* ed E. N. Zalta. Available online at: http://plato.stanford.edu/archives/spr2015/entries/structure-scientific-theories/

Yoda, K., Kira, T., Ogawa, H., and Hozumi, K. (1963). Self thinning in overcrowded pure stands under cultivated and natural conditions. *J. Biol. Osaka City Univ.* 14, 107–129.

Zamer, W. E., and Scheiner, S. M. (2014). A conceptual framework for organismal biology: linking theories, models, and data. *Integr. Comp. Biol.* 54, 736–756. doi: 10.1093/icb/icu075

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Scheiner and Holt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.