Ann. Zool. Fennici 39: 87–98 ISSN 0003-455X
Helsinki 14 June 2002

Are ecology and evolutionary biology "soft" sciences?

Massimo Pigliucci

Departments of Botany, Ecology & Evolutionary Biology, and Philosophy, University of Tennessee, Knoxville, TN 37996-1100, USA (e-mail: pigliucci @utk.edu)

Received 8 April 2002, accepted 25 April 2002

Pigliucci, M. 2002: Are ecology and evolutionary biology "soft" sciences? — *Ann. Zool. Fennici* 39: 87–98.

Research in ecology and evolutionary biology (evo-eco) often tries to emulate the "hard" sciences such as physics and chemistry, but to many of its practitioners feels more like the "soft" sciences of psychology and sociology. I argue that this schizophrenic attitude is the result of lack of appreciation of the full consequences of the peculiarity of the evo-eco sciences as lying in between a-historical disciplines such as physics and completely historical ones as like paleontology. Furthermore, evo-eco researchers have gotten stuck on mathematically appealing but philosophically simplistic concepts such as null hypotheses and *p*-values defined according to the frequentist approach in statistics, with the consequence of having been unable to fully embrace the complexity and subtlety of the problems with which ecologists and evolutionary biologists deal with. I review and discuss some literature in ecology, philosophy of science and psychology to show that a more critical methodological attitude can be liberating for the evo-eco scientist and can lead to a more fecund and enjoyable practice of ecology and evolutionary biology. With this aim, I briefly cover concepts such as the method of multiple hypotheses, Bayesian analysis, and strong inference.

"Eliminate all other factors, and the one which remains must be the truth."
(Sherlock Holmes, in Arthur Conan Doyle's *The Sign of the Four*)

A recurring complaint

Are ecology and evolutionary biology making progress, and if so, in what sense? On the face of it, the question may seem nonsensical. Just consider the number of papers on all sorts of evolutionary and ecological questions that are been published at a steady rhythm and one cannot have doubts about the fact that progress is being made. And yet, one also cannot avoid the nagging feeling that what we see in most published papers is the accumulation of new information, not necessarily progress in the sense of a conceptual understanding of the objects of study. The latter, after all, doesn't necessarily follow from the former.

I contend that little conceptual (as opposed to special problem-solving) progress has been made in what I will refer to as the evo-eco disciplines when compared to physics (the queen of the so-called "hard sciences") or even (but to a lesser extent) other biological fields such as molecular biology. Furthermore, I will argue that this is nothing to be ashamed of because it is at least in part the result of the very nature of ecology and evolutionary biology and of other "soft" sciences such as psychology and sociology. Of course, much depends on what one means by "progress" and how this is measured. However, as a practicing evolutionary ecologist I have often shared my frustration with too many colleagues about the fact that our fields don't seem to be going anywhere in particular for the problem to be only imagined. As Michael Turelli, a leading mathematical biologist, put it at the 2001 evolution meetings, "questions in our field are never settled, they simply go in and out of fashion". If you sympathize with this sentiment, read on.

Before considering why progress in evo-eco seems slow, one should of course be prepared to answer the more general question: is there progress in science at all? Again, this is a much more thorny issue than most scientists realize, and it requires familiarity with the philosophical concepts of truth, discovery, induction, and the like. I will not attempt to expound on these matters here, instead giving the reader my answer and some references to dig deeper. While clearly science cannot arrive at "the truth," whatever that is, it is equally clear that it has made progress in understanding the natural world (Hull 2000, Kitcher 1995, Maxwell 1979). In fact, somewhat ironically, Kitcher (Hull 2000) has used evolutionary theory as a paradigm of progress in science. Darwin's theory of evolution was "adopted on the basis of compelling reasons," and it was not simply the replacement of a theory with another, but a transition involving multi-faceted consequences on the practice of biology. It indubitably led to cumulative theoretical and empirical knowledge in the biological sciences.

The question that I wish to address here is more limited in scope and more focused on sub-theories within evo-eco: does research in evo-eco resemble more the hard sciences such as physics and chemistry, or the soft ones like psychology and sociology? This now familiar distinction was introduced by Windelband (1894/1980) in his *History and Natural Science*: what he called "nomothetic" knowledge is the one sought by most natural sciences and it consists in the discovery of general laws in order to understand and eventually master nature; what Windelband referred to as "idiographic" knowledge is typical of historical sciences and relies on descriptions of individual and unique aspects of reality, the main aim of which is reconstruction of events using a coherent narrative.

Some biologists, ironically from the ranks of paleontology – the most historically contingent of biological disciplines – have made a conscious but spotty effort to move our field as close as possible to the nomothetic "ideal" (Gould 1980, Raup & Gould 1974). Some further discussion of the potential and limits of the idea of evo-eco as hard science has been carried out since then, mostly by ecologists. The following brief discussion is not meant to be comprehensive, but only to assure the reader that I am not making this up. Plenty of serious researchers in our discipline have raised similar problems before, every time only to be largely ignored by the everyday torrent of puzzle-solving scientific papers that characterize most of what philosopher Thomas Kuhn (1970) called "normal science".

One of the loudest salvos was fired back in 1982 by none other than Ernst Mayr (1982). In a section of his *The Growth of Biological Thought* entitled "Laws in the physical and biological sciences" Mayr comes to several of the conclusions I discuss in more detail below, and in particular to the idea that evolutionary biology, being an inherently historical science, has by nature to follow a very different modus operandi than physics and the other "hard" sciences. As we shall see, I depart from Mayr when he insists that the focus on concepts is the key to understanding evo-eco.

A comprehensive analysis of the historical development of theories and concepts in ecology has been presented by McIntosh (1985) in his *The Background of Ecology: Concept and Theory*, which is a good starting point to understand the current status of this discussion. Peters' (1991) *A Critique of Ecology* contains reflections on the epistemology of ecological theory that actually apply to most of the scientific enterprise at large. Peters makes clear that an in-depth analysis of ecological statements from the point of view of the information they carry reveals troubles with the whole discipline (how many of us have jokingly told our students in introductory classes that "ecology is the elucidation of the obvious"?). Peters concludes that ecology should re-focus on simple questions of fact and observation, especially those of general relevance to science and society. This also happens to be my prescription, though reached by different means than Peters'.

From a different perspective, Lawton (1999) questions in what sense there are laws in ecology, concluding that while ecologists can produce generalized formulations based on observable tendencies, there is no place in ecology for laws in the sense meant in physics, i.e., statements about the natural world that are universally true. As a corollary, Lawton then suggests that ecologists should pay less attention to the "middle ground" of community ecology (an inflammatory statement, to be sure) and rely less on reductionism and experimental manipulation. As we shall see, my starting point below is similar, but I reach quite different conclusions.

Additional soul searching has been attempted by Shrader-Frechette and McCoy (1994) with their *Method in Ecology: Strategies for Conservation*, which while focusing on the practical aspect of conservation biology, has much to offer to evo-eco scientists in general. These authors begin by discussing the shaky grounds of such key concepts as "community" and "stability" and conclude that general and predictive theories in ecology are simply impossible. Their recipe for success falls along the lines of Peters' suggestions: what they call the "case study method" amounts to a serious focus on what ecology is (arguably) all about, natural history, good autoecology, and the clearest possible definition of hypotheses. Their discussion of the scientific vs. ethical implications of an emphasis on type I errors is interesting, though it suffers from a lack of reference to the obvious alternative to classical hypothesis testing that will be discussed below: Bayesian analysis. In a similar context, my discussion below of the tyranny of null hypotheses is particularly germane to the long, and still ongoing, controversy in ecology about the usefulness and relevance of "null models" (e.g., Gotelli & Graves 1996, Hubbell 2001).

These recurring self-examinations notwithstanding, most of us are much more preoccupied with carrying out business as usual, relegating questions of philosophy of science to coffee breaks and beer outings. Similar discussions, with similarly low impact on the everyday practice of science, have taken place for some time within the soft science par excellence, psychology, and the results are worth considering in some detail given what I think are very close parallels with evo-eco research.

Meehl (1978) has been perhaps one of the sharpest critics of the possibility of general theoretical progress in the soft sciences. His remark that "most theories in psychology never die, they slowly fade away" closely resembles Turelli's comment mentioned above. He went on to predict that we will probably never have a substantive general theory in personality or social psychology. Gergen (1973) went further to suggest that if the events of interest to psychologists are capricious, the discipline should be replaced by the equivalent of natural history (similar to Shrader-Frechette and McCoy's take on ecology discussed above), because the continued attempt to build general laws of social behavior may be misdirected. Since the study of social psychology is primarily an historical undertaking, he suggested that it would be best to think in terms of a continuum of historical durability of our empirical and theoretical findings, with phenomena highly susceptible to historical influence at one extreme (mostly psychology, and some natural sciences) and the more stable processes at the other (mostly chemistry and physics). For similar reasons, which I will examine below, it might be more difficult than we thought to achieve a general theory of ecology that cuts across population, community and ecosystem levels. And as far as evolutionary theory is concerned, there has been no conceptual unification of its various subdisciplines following the synthesis of the 1940s, and despite recent calls for such unification to which I participated (Schlichting & Pigliucci 1998), evo-eco biology currently looks more like a badly tossed salad than a melting pot of conceptual understanding.

So, what's the problem?

If any of the above should be of concern to evolutionary biologists and ecologists we need ask ourselves what the roots of the problem might possibly be. As we shall see, the answers identified by psychologists resonate with some of the well-known problems that plague our field as well. Box 1 lists a subset of the difficulties highlighted by Meehl in psychology that are relevant to evo-eco.

Some of the same points have been discussed by Cronbach (1975), who emphasized that the presence of higher-order interactions poses strict limitations to our understanding of complex phenomena. The usual rebuttal offered by classically quantitative-oriented biologists (and psychologists) is that interactions are not that important because they explain a small portion of the total phenomenological variance. This is misleading for a variety of reasons. First, as pointed out by Lewontin (1974) long ago the general linear model we use for most of our statistical analyses is biased in favor of main effects and against interactions. Second, implicit in the rebuttal is the fallacy that small variance implies lesser "importance". Yet, evolutionary biologists should know that small effects can be of overwhelming importance, as in the case of small but sustained selection pressures, or of mutations of small effect that provide the long-term fuel for evolutionary change.

A particularly simple but striking example (not from evo-eco, where the point is exactly that it is difficult to find such simple and striking situations) will make the point (Cronbach 1975). After the National Institutes of Health refurbished a lab for the study of how animals metabolize drugs, mice that used to sleep 35 minutes after injection of hexobarbital woke up after only 16 minutes. After a careful analysis,

Box 1. A list of problems and conceptual issues common to psychological and evo-eco research, from a subset in Meehl (1978).

1. Difficulties of slicing up the phenotype into meaningful intervals identified by causally relevant attributes.
2. Difficulty in achieving an adequate classification and sampling of environments and situations.
3. Difficulty in choosing appropriate scales of measurement and transformations.
4. Individual differences (e.g., genotypic variation, reaction norms).
5. Polygenic heredity of traits of interest.
6. Divergent causality (non-linearity), where differences in the exact character of the initial conditions are amplified over the long run.
7. Properties and relations (i.e., contingency) that make the study of living organisms rather more similar to such disciplines as history, archeology and geology.
8. Unknown critical events; sometimes these are observable events that were not actually observed, such as a demographic crash.
9. "Nuisance" variables: a non-negligible class of variables that are not random but systematic, exert a sizable influence, and are themselves also sizably influenced by other variables. Biologists refer to these as higher-order interactions.
10. Feedback loops and autocatalytic processes, the complexities of which are refractory to quantitative decomposition.
11. Sheer number of variables.
12. Limited correspondence between the results of lab and field experiments.
13. Inconsistency of results across labs.
14. Importance of exceptions and outliers.

investigators found out that the red-cedar bedding of the new cages made the difference, stepping up the activity of several enzymes that metabolize hexobarbital. The solution of the mystery required an explicit search for apparently insignificant interactions, one that was successful only because of the confined setting and the very limited number of potential factors. As Cronbach put it, “Once we attend to interactions, we enter a hall of mirrors that extends to infinity.” Any ecologist or evolutionary biologist who has ever worked with complex systems will feel déjà vu all over again.

Cronbach also realized that to get to the bottom of higher level interactions would require sample sizes that are simply logistically impossible, thereby setting an upper limit to what we can do given current approaches. He also highlighted a problem that is very well known to evo-eco researchers, but that he thought should be approached positively as an incentive to further research, as opposed to being minimized or even worse completely ignored. It is hardly the norm that results from field and laboratory research match. Worse yet, results obtained in different laboratories may not correspond either. Recently, there has been a number of papers in evolutionary biology confirming and quantifying this phenomenon (Ackermann *et al*. 2001, Hoffmann *et al*. 2001, Matos *et al*. 2000, Sgró & Partridge 2000). Should we then just throw our hands up and run for cover? No, because inconsistencies have causes as well. While some of these causes may not be of theoretical interest (such as experimental error, for example), others must lie in higher order interactions that can — with a subtle investigative work (as opposed to brute force, large experiments) — be identified and dissected. It is not by chance that a recent book devoted to a discussion of the most sophisticated methods in ecological analysis was entitled *The Ecological Detective* (Hilborn & Mangel 1997) and that some papers published in the main evolutionary literature explicitly use the detective metaphor to suggest the author’s approach to the problem at hand (Wills 1995).

A further problem that deserves a brief mention is the neglect that a classical quantitative approach to biological research has implied for the exceptions and the “outliers”. As it has been occasionally pointed out (Levin 1995, Lewontin 1966), our obsessive focus on means and other general descriptors of population behaviors automatically leads us to ignore exceptional behaviors or phenotypes as “anomalous” and “nonrepresentative”. But again, there is a real possibility that at least some of these anomalies actually fuel a significant amount of evolutionary change, or can provide ecological stepping stones that connect different populations. By concentrating exclusively on the average, do we risk ignoring crucial individual differences?

A possible solution? Strong inference

A problem similar to the one I am trying to bring into focus here was noted back in 1964 by Platt (1964). He asked himself why certain fields of research were making rapid progress and others were not, even within physics or biology themselves. His suggestion was that differential training and historical-cultural effects have led practitioners of some disciplines away from the application of what he called “strong inference”. This can be described with the following sequence:

1. Devise (several, not just two) alternative hypotheses.
2. Devise a crucial experiment(s), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses.
3. Carry out experiment(s) so as to get as clean results as possible.
4. Recycle the procedure, make sub-hypotheses or sequential hypotheses to refine the remaining possibilities.

He described Francis Crick’s lab at the time, in which a blackboard was always covered with “logical trees” of alternative hypotheses and students were excited at the prospect of discussing which experiments might cut some of the branches off the tree. This is of course also philosopher of science Karl Popper’s (1968) idea that the best way to proceed is by eliminating as many wrong answers as possible. Science is a procedure based on falsification more than on confirmation.

If this sounds familiar, it should, since strong inference is an elaboration on the basic proposal for a method of scientific investigation laid out by Francis Bacon (1620/2000) in his *New Organon*. More recently, Kitcher (1995) has also emphasized the central role of eliminative induction. This addresses the so-called problem of under-determination of theories by data (the fact that the same data can be explained by more than one theory: Okasha 2000, Shipley 2000). As Platt puts it: "When a group of hypotheses is at odds with some observational or experimental report, it is crucial to explore the explanatory losses that would occur if some members of the inconsistent set were eliminated. Making up the deficiency may involve new contradictions or new losses, generating a tree-like structure of possible adjustments to the corpus of beliefs." Furthermore, as Platt cunningly observed so many decades ago, getting into the habit of considering more than two possibilities (your favorite hypothesis and the "null" hypothesis) minimizes the always-present danger of falling so much in love with your theory that you become blinded to the actual verdict of the evidence (and most clearly put forward by T.C. Chamberlain [Chamberlain 1897]; reiterated throughout Hilborn and Mangel's Ecological Detective [Hilborn & Mangel 1997]; a point also made by Monod [Monod 1971]). Box 2 summarizes Platt's complaints about the sorry state of some scientific disciplines, which again I think closely resemble a rather common sentiment among ecologists and evolutionary biologists.

Platt's strong inference, however, is no panacea for the problems of evo-eco. While we can certainly make use of many of Platt's (and Bacon's) suggestions, it is not by chance that Platt's example of a successful application of the method was Francis Crick's lab: molecular biology is the least (though by no means completely lacking) historical of biological disciplines, the one most closely resembling physics and chemistry, and therefore, the one in which logical trees of sharply differentiated hypotheses are a productive algorithm. For psychology as well as evolutionary biology and ecology, things are a bit more complex, and we shall see that a more nuanced approach mirroring the philosophy embedded in Bayesian analyses is likely to bring us further.

Hard vs. soft science: what's the difference?

Despite the fact that things would surely improve if we were to stick to the principle of strong inference, the situation is just not that simple. There are other fundamental differences between hard and soft sciences, which we should be cognizant of in order to avoid chan-

Box 2. Platt's (1964) complaints about some bad practice in the natural sciences.

1. There are two kinds of biologists: those who are looking to see if there is one thing that can be understood, and those who keep saying it is very complicated and nothing can be understood.
2. Scientists become method- rather than problem-oriented. Stop doing experiments for a while and think.
3. "How many of us write down our alternatives and crucial experiments every day, focusing on the exclusion of a hypothesis?"
4. Small studies add another brick to the temple of science. But most such bricks just lie around the brickyard. They become substitutes for thinking, "a sad waste of intelligence in a research laboratory."
5. "We substitute correlations for causal studies, and physical equations for organic reasoning. Measurements and equations are supposed to sharpen thinking, but, in my observation, they more often tend to make the thinking non causal and fuzzy. They tend to become the object of scientific manipulation instead of auxiliary tests of crucial inferences."
6. THE crucial question to ask after a seminar: Sir, what experiment(s) could disprove your hypothesis? or: Sir, what hypothesis does your experiment disprove?

neling our efforts into frustrating dead ends. In fact, within biology, the modern distinction between genetics/molecular biology on one hand (hard sciences in a sense close to that of chemistry and physics) and evolutionary biology and ecology on the other was evident from the beginning. Genetics was born when Mendel switched from natural history observations to rigorous quantification and controlled experiments employing the strong inference approach. But the milestone in evolutionary biology was Darwin's *Origin of Species*, which includes little or nothing in the way of calculations and refers to few experiments. Instead, the latter is a masterpiece of detective work, putting the pieces of a complicated puzzle together one by one.

One of the major differences between the two kinds of endeavor is that in evo-eco sciences (as Cronbach already noticed for psychology: Cronbach 1975) the context in which research is done changes often, unlike in physics. An atom is an essentially a-historical object, so that it doesn't matter when and where you split it, you will get the same results. Biological organisms (and higher-level groupings such as populations and communities), on the other hand, are inherently and inextricably the outcome of many historical events (evolution, community assembly). This does not mean that biological research is hopeless, but it does imply that generalizations in evo-eco have short half-lives, so to speak. To be sure, things aren't quite that easy in physics either. While most people refer to the first part of Newton's *Principia* as the quintessential example of how hard sciences proceed, they also tend to neglect the second part, which deals with the complications of real — as opposed to ideal — bodies. Still today, the theory of nonlinear dynamics has shown why the three-body problem (the calculation of the exact positions of three celestial bodies orbiting around each other) is insoluble: the system is highly sensitive to initial conditions, so that even if it is entirely deterministic a precise solution valid at any particular moment is not obtainable. And we are talking about a very simple situation with no historicity to deal with. Imagine how difficult is to achieve a realistic mathematical treatment of objects as complex as populations or communities!

A more general point can be made concerning the complexity of evo-eco research that involves an understanding of basic philosophy of science. As we have seen, Popper (1968) suggested that the distinction between science and pseudoscience (the so-called "demarcation problem") can be drawn on the basis of the criterion of falsifiability. When a theory makes a risky (i.e., non-trivial) prediction, if it does not square with the empirical data the theory has been falsified. If a hypothesis is unfalsifiable (i.e., there is no way in principle to disprove it), then it is not scientific (which doesn't mean it's not true).

Lakatos (1977) understood that this sort of falsificationism is too simplistic: often we don't reject a previously supported hypothesis outright simply because the empirical results of one experiment don't fit it. Instead, we look first for possible experimental errors or systematic biases, and then we broaden our search to question the assumptions on which the hypothesis itself rested. So, empirical results cannot falsify a hypothesis, only the ensemble of hypothesis-assumptions-methods (the latter two components usually termed "auxiliaries"), which makes a strict application of the strong inference approach rather problematic. The problem is that in psychology (and evo-eco) the range of research circumstances in which auxiliary hypotheses are knotty is greater than in the exact sciences or in some but not all of the biological sciences (such as molecular biology). It is, therefore, hard to fulfill the Popperian requirement of stating beforehand what counts as a strong falsifier. In other words, some of our hypotheses may come close to pseudo-science, the just-so stories of some (but not all) sociobiologists and evolutionary psychologists come to mind as good examples.

There are however some myths concerning the difference between hard and soft sciences which need to be exploded because they too stand in the way of our understanding of science as a process. One of these myths is that physics, for example, is a paragon of consistency in its results, while soft sciences are so frustrating because things tend to be different between different experiments or situations. Meehl (1978) has tested this prediction by conducting an extensive meta-analysis of results from several subfields of physics and psychology. It turns out that there is no large difference between the consistency of results from the social and physical sciences. The idea that results in physics are strikingly consistent and in psychology strikingly inconsistent is simply not supported by the empirical evidence. Therefore, the results of social science research are reasonably empirically (though not necessarily conceptually) cumulative when compared with the results from physics. Might the same be true for evo-eco research? What Meehl also found was that the main difference between physical and social sciences' results is to be found in the fact that the first are much more *accurate* (where accuracy is the ratio between the measured value and its standard error). However, this measure might mean very little given some astounding data reported in his paper. Meehl shows a figure with two curves relating temperature to thermal conductivity of gadolinium. The accuracy of the first curve was stated as within 1% and that of the second one as 0.5%. Yet, the two curves differed from each other by up to 500%! Are our beloved standard errors really measures of how far our estimates are from the true mean of the population, or are we fooling ourselves in the false security of "hard" numbers?

Against the null hypothesis

A more general criticism of some soft sciences that could be applied also to ecology and evolutionary biology emerges from the works of Cronbach, Meehl, and Platt: simply put, we may be relying too much on classical statistics. Meehl goes as far as saying that Sir Ronald Fisher has "befuddled us, mesmerized us, and led us down the primrose path" to which Cronbach echoed "the time has come to exorcise the null hypothesis". These are harsh words, but the arguments should resound with anybody who has any experience of real research in the social sciences and — I maintain — in evo-eco. (For a delightfully ironic and clear piece on the same subject, *see* Cohen 1994's "The Earth is round ($p < 0.05$)").

The idea is that most of our predictions, especially when expressed as null hypotheses and their rejection, are quite trivial. In the same way as one should not be impressed by a statement that on average November is going to be a colder month than September, no matter how many asterisks accompany it, we should not be overly confident in the way we formulate null hypotheses. Most of our H_0s are, in fact, obviously false or likely to be false before we even start the experiment, which means that rejecting them does not advance our understanding very much. Furthermore, we often behave as if there were only two alternatives, triumphantly claiming that our alternative hypothesis is supported by the rejection of the null. But things are obviously not that simple. There are always many other possibilities, and we need to get into the habit of stating multiple hypotheses as precisely as possible because the alternative(s) to the null just don't win by default. Moreover, we need to give up the idea that individual experiments are going to accomplish much. Given the complexity, puzzle-like structure, of the problems we investigate, each experiment will be at best able to diminish the likelihood of one or two possibilities, certainly not to confirm our favorite hypothesis beyond reasonable doubt. The problem is that evo-eco is a high-information field characterized by usually low-information empirical studies, and we just have to learn how to live with this.

Meehl (1978) again points out circumstances that are all too familiar to evo-eco researchers. For example, he describes a typical situation where some tests (say, seven out of ten) favor the hypothesis and some (three out of ten) don't, and the author of the hypothetical paper concludes that "further research is needed to explain the discrepancies," without realizing that the non-supportive tests do much more damage to the theory than the support provided by the positive tests. Meehl goes on to suggest that when an author tries to 'make theoretical sense' out of such a table of favorable and adverse significance test results, what she is actually engaged in, willy-nilly or unwittingly, is meaningless substantive constructions on the properties of the statistical power function, and almost nothing else. A theory that has seven tests in favor and three against is *not* in good shape.

There is a fundamental and often neglected difference between a substantive theory and a statistical hypothesis: the latter is a very restricted operational concept that is only indirectly connected to the former. But it is the substantive theory that we most care about.

Meehl concludes that we should ask ourselves what kind of inferred entity construction we want and how it could generate the sorts of intellectual surprises that are typical of "invisible hand theories". Because of Fisher's legacy and a long history of over-application of statistics we may have thrown in the towel and abandoned hope of concocting substantive theories that will generate stronger consequences than merely X differs from Y. At the very least we should be able to predict the order of numerical values or the rank of the first-order numerical differences. A minimalist theory should be able to generate at least a certain function form, such as a graph with a given shape and number of peaks. This is rarely the case in evolutionary biology and ecology.

The method of multiple hypothesis and Bayesian inference

The reader will have guessed by now that an obvious alternative to the classical way of proceeding is provided by so-called Bayesian analysis. I cannot provide here a review of this field, its underpinnings and its usefulness in biological research (but *see*, for example: Hilborn & Mangel 1997, Howson & Urbach 1991, Huelsenbeck *et al*. 2000, Jefferys & Berger 1992, Malakoff 1999, Rudge 1998, Shoemaker *et al*. 1999). However, what is of interest to our discussion is Bayesianism as a *philosophical* model for doing science, rather than the actual nitty-gritty (and quite complex) details of how to implement it in real research applications.

The fundamental theorem of Bayesian statistics (which was *the* way to do statistics until Fisher took over in the early 20th century) is:

$$P(H_i \mid D) = \frac{P(D \mid H_i) \times P(H_i)}{P(D \mid H_i) \times P(H_i) + \ldots + P(D \mid H_n) \times P(H_n)}$$

where we are considering a series of hypotheses $i \ldots n$, D stands for the observed data at a given moment during the research project, H for a certain hypothesis, P for a probability, and $|$ is the conditional probability operator. What the equation says is that the probability of a certain hypothesis given the available data (the so-called "posterior" probability) is the ratio of two quantities: the product of the probability of the data given the hypothesis in question (the so-called likelihood of the data) multiplied by the a priori probability of that hypothesis (the so-called "prior") at the numerator, and the same quantity summed over all considered hypotheses at the denominator (sometimes referred to as the total probability of the data).

It should be obvious why a Bayesian approach directly addresses many of the problems we have discussed so far. For starter, it embodies the idea of multiple competing hypotheses in the definition of the denominator of Bayes' rule: there are no privileged ("null") hypotheses, but only a fair competition among stated alternatives. Second, the question is posed in more sensible terms then with the classical approach: we are asking what is the probability of a certain hypothesis given the data, not the other way around (which is what the Fisherian method asks). Third, the concept of probability in Bayesian analyses is different from the standard one and more appropriate for scientific research: a probability here is not the frequency with which a certain outcome would occur if we repeated the experiment *n* times (as it is in the Fisherian, also known as "frequentist" approach), but an estimate of the degree of belief (as in likelihood, not blind faith) we are entitled to attach to a given hypothesis because of what we know of the problem (including the data collected to answer the question). Fourth, Bayesian analysis — unlike the classical approach — takes into account what we knew of the problem before starting our experiment (in the form of the priors introduced above), which makes sense because we never start a project with our mind set to the state of a *tabula rasa* and it is a good idea to quantify as much as possible our prior ideas about the problem at hand. Finally, a Bayesian framework makes it very difficult to think in terms of either naïve falsificationism or naïve confirmationism: we cannot get zero (complete falsification) or one (complete confirmation) as posteriors for any given hypothesis. All we can reasonably hope for is to see our belief in some hypotheses (ideally just one) go significantly up after an experiment and our belief in as many other hypotheses as possible go down accordingly. Indeed, this change can be measured independently for any given hypothesis, which gives the researcher an estimate of how informative the experiment was in respect to each hypothesis.

While Bayesianism is itself no panacea for either statistical inference or scientific methodology (for example, it is tricky to set priors reasonably, and a long discussion has been going on about the difference between "objective" and "subjective" priors), the reason it strikes me as the best model on the market (itself a Bayesian statement, incidentally) is that it resonates very well with the actual practice of science as I have experienced it in decades of academic research. We think like Bayesians, even though we often constrain our papers within a Fisherian straight jacket.

The compromise: evolution and ecology as nomothetic and idiographic sciences

As Francis Bacon noted (1620/2000), it is no good to engage only in the *pars destruens* (i.e., the negative criticism) of an argument, unless one has a *pars construens* (positive suggestions) to add to it. Despite the criticisms and warnings that I discussed above, I do of course think that evolutionary biology and ecology are (mostly) real sciences and that progress has indeed occurred and still does. The real questions are: where do they lie between the two extremes of nomothetic and idiographic sciences, and what does this mean for our everyday practice as researchers?

First, it seems to me undeniable that ecology and evolutionary biology are partly historical and partly a-historical sciences. A particularly clear example (among many) comes from a recent study of changes in variance-covariance (**G**) matrices in *Drosophila*. Phillips and collaborators (2001) have investigated the effect of genetic drift on the similarity of **G** matrices in 52 independently derived inbred lines of fruit flies when compared to outbred controls. They found that the *average* results were in perfect agreement with the theoretical predictions: drift alters only the size but not the shape of **G** matrices. However, they also found that it is impossible to predict the actual shape of any given **G** due to a large amount of individual variation among the inbred lines. Since this variation would translate in significantly different evolutionary trajectories, we are in a paradigmatic example of success of average predictions (which tend to be a-historical) and abysmal failure of specific predictions (which are markedly influenced by history). This is exactly what philosopher of science John Dupré (1993) predicted based on his theory of non-reductionism in the sciences: in the case of complex systems, theoretical reduction can explain the boundaries of variation of the observed phenomena, but will fail to tell us what exactly is going to happen in any specific case (on the general problem of the basis of biological explanations *see* Rosenberg 2001). This may be a general property of the world (Kauffman 1993) with which we simply have to be able to live and even take advantage of.

As a consequence of the partial historicity of evo-eco research, our sciences are really somewhere between the nomothetic extreme of fundamental physics and the idiographic end of, say, paleontology, anthropology, and some of the social sciences (notice that this does not mean that historical disciplines cannot make testable predictions, only that it is more difficult to do so when history is a factor). This means that we should drop our physics envy and work more like puzzle-solvers, adopting a healthy balance of observation, experimentation, and mathematical theorizing. As of now, the balance seems to be much to much at the expense of observation, branded as a somewhat second-grade activity reserved for natural historians (a term that has itself unofficially become a slur within certain academic circles). We should think of ourselves as Sherlock Holmes, not as Isaac Newton. Far from being demeaning, it can be a liberating feeling (for example, see a paper on data analysis as detective work by a statistician of the caliber of Tukey: Tukey 1969).

Another consequence of a new attitude toward the complexities of evo-eco research is that our *modus operandi* should embrace the concept of consilience. The word "consilience," made familiar to biologists through a book by E.O. Wilson (1998) was introduced by English philosopher William Whewell (1840) to explain the phenomenon that often pieces of evidence from disparate sources "jump together" toward a common explanation, what he termed a consilience of induction (in psychological terms this is a Gestaltian experience, and it may be profoundly linked to the way our brain works and interprets reality: Gazzaniga 2000). While Whewell's work on induction was overshadowed by the later contribution of John Stuart Mill, he was the first philosopher to resurrect the importance of induction in science. In his seminal paper, Whewell stated that "Accordingly the cases in which inductions from classes of facts altogether different have thus jumped together, belong only to the best established theories which the history of science contains. And, as I shall have occasion to refer to this particular feature in their evidence, I will take the liberty of describing it by a particular phrase; and will term it the Consilience of Inductions." In other words, we can reach (provisional) conclusions in complex matters by a sort of triangulation in logical space, when different types of evidence point toward the same answer. And the more stringent the triangulation, the more likely (but never certainly) we can pinpoint the "culprit".

Perhaps the most delicate consequence of realizing what kind of science evo-eco really is concerns the impact that such realization should have on funding and publishing priorities. During recent times we have witnessed a marked movement toward more "hard" disciplines and topics of scientific research, such as genomics, as well as an equally clear preference for mathematical and statistical approaches to evo-eco questions. While this is all certainly very valuable and should be continued, it is time to pause and realize the limitations intrinsic in these choices. As editors, reviewers and officers of societies and funding agencies we should ask what kinds of questions are best pursued and how in order to make choices that are influenced by a solid philosophy of science rather than by fashion or novelty.

The warnings mentioned above, from a variety of authors, concerning over-reliance on statistical testing, simplistic formulation of hypotheses, cavalier interpretations of results, and a tendency to substitute technique for thinking are all topics that are very familiar to philosophers of science. Alas, most of us practicing scientists just plug ahead with our research, confining the tough questions to evening discussions over a beer with our graduate students. And yet, it does pay to occasionally look at the forest instead of individual trees, and search for the best road through the vegetation by adopting a bird's eyeview. Contrary to what is maintained by some "hard" scientists (Weinberg 1992), philosophers of science could be extremely useful to the practical scientist, if only we would stop a moment to listen to what they are saying (Wilkins 2001).

Acknowledgments

I wish to thank Jonathan Kaplan, William Provine and Marlene Zuk for careful comments on previous drafts of this paper and Elliott Sober for a critical discussion of Bayesianism. This work was partially supported by NSF grant DEB-0089493.

References

Ackermann, M., Bijlsma, R., James, A. C., Partridge, L., Zwaan, B. J. & Stearns, S. C. 2001: Effects of essay conditions in life history experiments with *Drosophila melanogaster*. — *Journal of Evolutionary Biology* 14: 199–209.

Bacon, F. 1620/2000: *Novum Organum*. — Cambridge University Press, Cambridge, England.

Chamberlain, T. C. 1897: The method of multiple working hypotheses. — *Science* 15: 92–96.

Cohen, J. 1994: The earth is round ($p < 0.05$). — *American Psychologist* 49: 997–1003.

Cronbach, L. J. 1975: Beyond the two disciplines of scientific psychology. — *American Psychologist* 30: 116–127.

Dupré, J. 1993: *The disorder of things: metaphysical foundations of the disunity of science*. — Harvard University Press, Cambridge, MA.

Gazzaniga, M. S. 2000: Cerebral specialization and interhemispheric communication. Does the corpus callosum enable the human condition? — *Brain* 123: 1293–1326.

Gergen, K. J. 1973: Social psychology as history. — *Journal of Personality and Social Psychology* 26: 309–320.

Gotelli, N. J. & Graves, G. R. 1996: *Null models in ecology*. — University of Chicago Press, Chicago.

Gould, S. J. 1980: The promise of paleobiology as a nomothetic, evolutionary discipline. — *Paleobiology* 6: 96–118.

Hilborn, R. & Mangel, M. 1997: *The ecological detective: confronting models with data*. — Princeton University Press, Princeton, NJ.

Hoffmann, A. A., Hallas, R., Sinclair, C. & Patridge, L. 2001: Rapid loss of stress resistance in *Drosophila melanogaster* under adaptation to laboratory culture. — *Evolution* 55: 436–438.

Howson, C. & Urbach, P. 1991: *Scientific reasoning: the Bayesian approach*. — Open Court, La Salle, IL.

Hubbell, S. P. 2001: *The unified neutral theory of biodiversity and biogeography*. — Princenton University Press, Princenton, NJ.

Huelsenbeck, J. P., Rannala, B. & Masly, J. P. 2000: Accomodating phylogenetic uncertainty in evolutionary studies. — *Science* 288: 2349–2350.

Hull, D. 2000: *Science & selection: essays on biological evolution & the philosophy of science*. — Cambridge University Press, New York, NY.

Jefferys, W. H. & Berger, J. O. 1992. Sharpening Ockam's razor on a Bayesian strop. — *American Scientist* 80: 64–72.

Kauffman, S. A. 1993: *The origins of order*. — Oxford University Press, New York.

Kitcher, P. 1995: *The advancement of science: science without legend, objectivity without illusions*. — Oxford University Press, New York, NY.

Kuhn, T. 1970: *The structure of scientific revolutions*. — University of Chicago Press, Chicago.

Lakatos, I. 1977: *The methodology of scientific research programmes*. — Cambridge University Press, Cambridge, England.

Lawton, J. H. 1999: Are there general laws in ecology? — *Oikos* 84: 177–192.

Levin, D. A. 1995: Plant outliers: an ecogenetic perspective. — *The American Naturalist* 145: 109–118.

Lewontin, R. C. 1966: Is nature probable or capricious? — *BioScience* January: 25–27.

Lewontin, R. C. 1974: The analysis of variance and the analysis of causes. — *American Journal of Human Genetics* 26: 400–411.

Malakoff, D. 1999: Bayes offers a 'new' way to make sense of numbers. — *Science* 286: 1460–1464.

Matos, M., Rose, M. R., Pité, M. T. R., Rego, C. & Avelar, T. 2000: Adaptation to the laboratory environment in *Drosophila subobscura*. — *Journal of Evolutionary Biology* 13: 9–19.

Maxwell, N. 1979: Induction, simplicity and scientific progress. — *Scientia* 114: 629–674.

Mayr, E. 1982: *The growth of biological thought: diversity, evolution, and inheritance*. — Belknap Press, Cambridge, MA.

McIntosh, R. P. 1985: *The background of ecology: concept and theory*. — Cambridge University Press, Cambridge, England.

Meehl, P. E. 1978: Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. — *Journal of Consulting and Clinical Psychology* 46: 806–834.

Monod, J. 1971: *Chance and necessity; an essay on the natural philosophy of modern biology*. — Knopf, New York, NY.

Okasha, S. 2000: The underdetermination of theory by data and the "Strong Programme" in the sociology of knowledge. — *International Studies in the Philosophy of Science* 14: 283–297.

Peters, R. H. 1991: *A critique for ecology*. — Cambridge University Press, Cambridge, England.

Phillips, P. C., Whitlock, M. C. & Fowler, K. 2001: Inbreeding changes the shape of the genetic covariance matrix in *Drosophila melanogaster*. — *Genetics* 158: 1137–1145.

Platt, J. R. 1964: Strong inference. — *Science* 146: 347–353.

Popper, K. R. 1968: *Conjectures and refutations; the growth of scientific knowledge*. — Harper & Row, New York, NY.

Raup, D. M. & Gould, S. J. 1974: Stochastic simulation and evolution of morphology — Towards a nomothetic paleontology. — *Systematic Zoology* 23: 305–322.

Rosenberg, A. 2001: How is biological explanation possible? — *British Journal for the Philosophy of Science* 52: 735–760.

Rudge, D. W. 1998: A Bayesian analysis of strategies in evolutionary biology. — *Perspectives on Science* 6: 341–360.

Schlichting, C. D. & Pigliucci, M. 1998: *Phenotypic evolution, a reaction norm perspective*. — Sinauer, Sunderland, MA.

Sgró, C. M. & Partridge, L. 2000: Evolutionary responses of the life history of wild-caught *Drosophila melanogaster* to two standard methods of laboratory culture. — *American Naturalist* 156: 341–353.

Shipley, B. 2000: *Cause and correlation in biology: a user's guide to path analysis, structural equations and causal Inference*. — Cambridge University Press, Cambridge, England.

Shoemaker, J. S., Painter, I. S. & Weir, B. S. 1999: Bayesian statistics in genetics: a guide for the uninitiated. — *Trends in Genetics* 15: 354–358.

Shrader-Frechette, K. S. & McCoy, E. D. 1994: *Method in ecology: strategies for conservation*. — Cambridge University Press, Cambridge, England.

Tukey, J. W. 1969: Analyzing data: sanctification or detective work? — *American Psychologist* 24: 83–91.

Weinberg, S. 1992: *Against philosophy*. *Dreams of a final theory*: 166–190 Pantheon, New York.

Whewell, W. 1840: *The philosophy of the inductive sciences, founded upon their history*. — J.W. Parker, London.

Wilkins, A. S. 2001: Why the philosophy of science actually does matter. — *BioEssays* 23: 1–2.

Wills, C. 1995: When did Eve live? An evolutionary detective story. — *Evolution* 49: 593–607.

Wilson, E. O. 1998: *Consilience: the unity of knowledge*. — Knopf : Distributed by Random House, New York, NY.

Windelband, W. 1894/1980: History and natural science. — *History and Theory* 19: 169–185.