Environmental Conservation **35** (3): 193–196 © 2008 Foundation for Environmental Conservation doi:10.1017/S0376892908004888

COMMENT

Making empirical progress in observational ecology

Epistemology, the theory of how knowledge arises, is an issue that ecologists and environmental scientists may not often consider in their work. However, beliefs about epistemology play a direct role in how they conduct research and, consciously or not, the conclusions made through analysis reflect some form of epistemology. I argue that the Popperian epistemology by which most environmental scientists are trained is generally incompatible with research on observational data and that observation-based studies can be improved by modifying Popper's notion of falsification.

The basic premise of scientific reasoning is that knowledge arises through observations that support some idea about how the natural world works (Feynman 1985). This 'knowledge that is known by experience' is the definition of empiricism, the method of scientific understanding. A scientist develops ideas about a subject of interest based on personal experience and proceeds to test the validity of their ideas through observation or experiment. This inclusive definition of scientific enquiry was, in some sense, constrained by Karl Popper (1902–1994), who crystallized scientific practice in the 20th century by asserting that scientific hypotheses cannot be proved, but only disproved (Popper 1934).

Popper made a strong distinction between inductive reasoning (things that are posited to be true based on repeated observation) and deductive reasoning (things that are described to be true by working from general principles to reach specific conclusions), stating that, while inductive arguments are sometimes useful in the practice of science, they have no place in its epistemology: 'the doctrine that the empirical sciences are reducible to sense perceptions, and thus to our experiences, is one which many accept as obvious beyond all question. However, this doctrine stands or falls with inductive logic, and is here rejected along with it' (Popper 1934).

Popper eliminated inductive arguments from the practice of science by requiring that scientific ideas be subject to falsification, where empirical scientific facts are restricted to ideas that can be subjected to critical tests of predicted outcomes. Only through a series of successive positive-outcome experiments can a hypothesis gain a measure of scientific support, always with the potential for failure or rejection in the next experiment ('a positive decision can only temporarily support the theory'; Popper 1934). For Popper, falsifiable equalled empirical.

The inferential power of falsification is familiar to most scientists; it is Popper's best known idea on the development of scientific reasoning and taught to every science student. Empirical testing of ideas thorough experimentation is the strongest possible form of scientific inference (a conclusion based on evidence and reasoning), whereby subjects are randomly assigned to control and treatment groups that differ by known quantities developed by the experimenter (Gelman *et al.* 2004). Because subjects are assigned randomly, the post-experiment differences between control and treatment groups are attributed directly to the effect of the experimental treatment. These conditions lead to strong inferences because the potential for unknown factors to influence the results is kept to the minimum extent possible (Hurlbert 1984; MacKenzie *et al.* 2006).

Yet for many important research questions, such as those relating to climate change, no experimental manipulations can be made and data are simply observed as having arisen from natural system dynamics (Underwood 2000; MacKenzie *et al.* 2006). Purely observation-based studies lie at the opposite end of the strength-of-inference scale from laboratory experiments, with field experiments (where researcher manipulations are made *in situ*) and natural experiments (where researchers have no input; for example extant marine protected areas) lying in between. Observational studies are frequently conducted at large scales (Carpenter 1990) or over long time periods (Carpenter 2002), where no two systems are the same and controls are often impossible to obtain (Loehle 1990; Hilborn & Ludwig 1993); frequently they will form a basis for subsequent experiments of lower-scaled processes (Underwood 2000). These studies include many kinds of ecological data, including those generated by disturbance, long-term monitoring, harvest records, space-for-time substitutions, gradient analysis and single-season observations over a wide survey area (for example Willis & Anderson 2003; Fukami & Wardle 2005). Because there are no treatments present, observational studies cannot guarantee that unknown factors have not unduly influenced the data observed and, as such, they constitute the weakest possible form of inference. They also differ from the Popperian model by being almost entirely inductive (Quinn & Dunham 1983).

The clear distinction that exists between Popper's deductive empiricism through falsification and the practice of inductive empiricism via observation may be evident to those familiar with the philosophy of science, but for many students and scientists conducting ecological research there remains the persistent notion that inductive studies must be subject to falsification if the results are to be considered legitimate. The need to fulfil this strongly held (and repeatedly taught) belief has, in turn, led to a lack of distinction among two sets of ideas, between deductive versus inductive reasoning and between Popperian testing versus the statistical testing of data (Table 1). The need to falsify has led environmental scientists conducting observational studies (Table 1, middle column) to

https://doi.org/10.1017/S0376892908004888 Published online by Cambridge University Press

Table 1 Contingency table between the logical form of a scientific study and two potential methods of inference.

Type of testing	*Logical form*	
	Inductive	*Deductive*
Popperian	No	Yes
Statistical	Yes	Yes

frequently rely on test statistics (Table 1, lower row) in the belief that they are achieving Popperian falsification of their ideas (Table 1, upper row). Although this concept has, perhaps unintentionally, been reinforced by leading authorities (Underwood 2000), rather than providing a case for falsification, it represents an abuse of test statistics that masks the empirically-weak nature of observational studies.

Null hypothesis testing was developed by Fisher (1925) for biologists to analyse 'numerical data accumulated in their own laboratories'. At the time, Fisher was concerned with developing methods for rigorous experimental protocols to isolate individual causes (Grace 2006) and, as quantities that could be easily calculated, hypothesis tests were ideal for evaluating the results of experimental studies where subjects were randomly allocated to experimental groups. For deductive experiments, hypothesis testing provided a direct link between objective statistical analysis and the objective testing criteria that Popper proposed. The *de facto* connection between these ideas has not frequently been discussed (Hilborn & Mangel 1997) yet they have become synonymous, whereby completing a hypothesis test means that a researcher is being empirical and doing experiments. But for environmental scientists who must employ observational data in their research, this is decidedly not what is happening; most often they are studying multiple interacting processes, none of which are (or can be) controlled in an experiment.

If observation-based studies are not experiments, how should ecologists approach the analysis of observed data? First they must be clear about what they are doing and stop treating their studies as experiments. Both the desire for Popperian falsification and a devotion to hypothesis tests must be abandoned if data are only observed. This requires a change in perspective that, once made, will direct attention to the quantities of interest in any good statistical data analysis, namely effect sizes and their relative uncertainty (Hobbs & Hilborn 2006). But if this eliminates Popperian falsification as a method of inference, how can observation-based studies be empirical?

The solution lies in the foundation of Popper's own theory, whereby scientific ideas must be routinely challenged to gain scientific support. In discarding inductive arguments, Popper was attempting to find a way for scientists to rigorously and objectively challenge their own ideas, ideas that, as humans, we have a strong tendency to defend beyond reason. Although Popper chose falsification as the method by which he believed objective criticism could develop, it is in fact the act of criticizing ideas with evidence (not falsification itself) that is the foundation of the scientific method. For ideas to be scientific, what is believed to be true must repeatedly stand up to counter-argument. This is in contrast to, for instance, spiritual or religious ideas that require only belief to attain validity. By viewing the 'testing' that Popper proposed as the act of 'challenging', alternative explanations become the primary means for validating a given idea. And if environmental scientists successively challenge their ideas with alternative explanations, then their work will readily fall within the empirical domain. How then can this approach be put into practice?

In observational studies, alternative explanations can often take the form of alternative models that can be statistically compared to provide evidence for one idea relative to others (Hilborn & Mangel 1997). These so-called multiple hypothesis methods have recently gained wide support (Hobbs & Hilborn 2006), in particular because testing an a priori set of potential models leads to stronger inferences than alternative approaches such as data mining (Burnham & Anderson 2002). For example, Vucetich *et al.* (2002) compared multiple models to evaluate the relative evidence for prey (moose, *Alces alces*) versus predator (wolf, *Canis lupus*) densities affecting predator kill rates, finding a greater level of support for predator effects than for prey effects. They clearly understood the role of alternative models for observation-based inference, pointing out that they had 'directly observed the influence of both prey and predator density on kill rates in an unmanipulated terrestrial system' (Vucetich *et al.* 2002). With no experiments present, such observation-based studies cannot rely on predicted outcomes for inferential support and the objective of analysis is then to determine the relative importance of, and interactions between, multiple processes that are reasonably suspected of operating (Quinn & Dunham 1983). This can happen at many levels, from parameters to sub-models, or for entire paradigms.

For example, the field of reef ecology has held a protracted argument about what processes determine the highly-diverse distribution of reef fishes observed throughout the tropics. Major processes include dispersal and recruitment (Doherty 2002, pp. 327–355), habitat structure (Galzin 1987), predation (Hixon 1991) and competition (Almany 2003). Many well-designed field experiments have been conducted over many years to demonstrate the existence of each of these phenomena, and few reef ecologists would now argue that a single mechanism is responsible for the observed diversity of reef fishes. Where much of the current debate lies however, is in determining the relative importance of these processes in real ecosystems, and it is here that inclusion of multiple hypotheses could serve reef ecologists well (Shima 1999).

A major applied problem is that reefs are under threat from disturbance on many fronts (Bellwood *et al.* 2004) and to predict future reef states requires some understanding of the relative roles played by potential diversity processes under changing environmental conditions (Wilson *et al.*

https://doi.org/10.1017/S0376892908004888 Published online by Cambridge University Press

2006). For example, after a strong coral bleaching event, habitat conditions might determine species diversity, while in pristine environments, recruitment could be the process of greatest importance; at this point no one can be sure. Ideally researchers could manipulate entire reefs to determine the effects of system-wide disturbance (*sensu* Simberloff & Wilson 1969), however the threatened status of most coral reefs make large-scale field experiments unfeasible; reef scientists must therefore rely on observed disturbance events to predict what will happen to disturbed reefs in the future. Given the opportunity, repeated observations from disturbed and undisturbed reefs used to estimate the relative effects of alternative diversity-generating processes would help to predict reef community states under variable disturbance conditions. Such an approach to inference would go a long way to improve the general applicability of observation-based studies on reefs, where the 'logic of many case studies' (Schrader-Frechette & McCoy 1994) should lead to better conclusions.

Although ecologists have considered many aspects of scientific inference previously (Simberloff 1983; Hurlbert 1984; Belovsky *et al*. 2004), they have skirted the fact that observation-based studies cannot rely on Popperian falsification. Despite the conceptual simplicity of the approach proposed here, evaluating the relative importance of multiple ideas is nonetheless a challenging task. Poorly conducted studies using variables that are biased to a particular result will not be fair arbiters of relative effects and it may be difficult (or impossible) to obtain data that support particular hypotheses; this has often been the case for recruitment data from reef fish (Doherty & Fowler 1994; James *et al*. 2002). However, in any properly-conducted study (experimental or observational) it is the responsibility of individual scientists to keep competing ideas in contention during project development (Underwood 2000). If research is undertaken with only one theory in mind then evidence for that theory is much more likely to be forthcoming; this is poor science in any instance. Scientists relying on observational data have the additional burden of operating without the use of experiments in an inferentially-weak discipline. To overcome these deficiencies researchers must be ready to challenge their hypotheses with ideas that may end up being more important than their own.

Acknowledgements

I would like to thank N.V.C Polunin for the invitation to write this article; M.J. Conroy and C.J. Fonnesbeck for numerous discussions regarding observation-based ecology; and L. Thomas, A. Edwards, J. Carlson, and R. McMaken for reviewing previous drafts. Particular thanks go to Trevor Willis and two anonymous reviewers for their critical analysis of the text. This work was funded by the National Science and Engineering Research Council of Canada and the National Research Council, USA.

References

Almany, G. (2003) Priority effects in coral reef fish communities. *Ecology* **84**(7): 1920–1935.

Bellwood, D., Hughes, T., Folke, C. & Nyström, M. (2004) Confronting the coral reef crisis. *Nature* **429**: 827–833.

Belovsky, G., Botkin, D., Crowl, T., Cummins, K., Franklin, J., Hunter Jr, M., Joern, A., Lindenmayer, D., MacMahon, J., Margules, C. & Scott, J. (2004) Ten suggestions to strengthen the science of ecology. *BioScience* **54**(4): 345–351.

Burnham, K. & Anderson, D. (2002) *Model Selection and Multimodel Inference: a Practial Information-theoretic Approach*. Second edition. New York, NY, USA: Springer-Verlag.

Carpenter, S. (1990) Large-scale perturbations: opportunities for innovation. *Ecology* **71**(6): 2038–2043.

Carpenter, S. (2002) Ecological futures: building an ecology of the long now. *Ecology* **83**(8): 2069–2083.

Doherty, P. (2002) *Coral Reef Fishes. Dynamics and Diversity in a Complex Ecosystem*. San Diego, CA, USA: Academic Press.

Doherty, P. & Fowler, T. (1994) An empirical test of recruitment limitation in a coral reef fish. *Science* **263**(5149): 935–939.

Feynman, R. (1985) *Surely You're Joking Mr Feynman*. New York, NY, USA: W.W. Norton and Sons.

Fisher, R. (1925) *Statistical Methods for Research Workers*. London, UK: Oliver and Boyd.

Fukami, T. & Wardle, D. (2005) Long-term ecological dynamics: reciprocal insights from natural and anthropogenic gradients. *Proceedings of the Royal Society of London B* **272**: 2105–2115.

Galzin, R. (1987) Structure of fish communities of French Polynesian coral reefs. II. Temporal scales. *Marine Ecology – Progress Series* **41**: 137–145.

Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2004) *Bayesian Data Analysis*. Second edition. New York, NY, USA: Chapman and Hall.

Grace, J. (2006) *Structural Equation Modelling and Natural Systems*. Cambridge, UK: Cambridge University Press.

Hilborn, R. & Ludwig, D. (1993) The limits of applied ecological research. *Ecological Applications* **3**(4): 550–552.

Hilborn, R. & Mangel, M. (1997) *The Ecological Detective*. Monographs in Population Biology, no. 28. Princeton, USA: Princeton University Press.

Hixon, M. (1991) Predation as a process structuring coral reef fish communities. In: *The Ecology of Fishes on Coral Reefs*, ed. P. Sale, pp. 475–508. San Diego, CA, USA: Academic Press.

Hobbs, N. & Hilborn, R. (2006) Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Ecological Applications* **16**(1): 5–19.

Hurlbert, S. (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**(2): 187–211.

James, M., Armsworth, P., Mason, L. & Bode, L. (2002) The structure of reef fish metapopulations: modelling larval dispersal and retention patterns. *Proceedings of the Royal Socitey of London B* **269**: 2079–2086.

Loehle, C. (1990) Philosophical tools: reply to Shrader-Frechette and McCoy. *Oikos* **58**(1):115–119.

MacKenzie, D., Nichols, J., Royle, J., Pollock, K., Hines, J. & Bailey, L. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. San Diego, CA, USA: Elsevier.

Popper, K. (1934) *Logik der Forschung*. Vienna, Austria: Mohr Siebeck.

https://doi.org/10.1017/S0376892908004888 Published online by Cambridge University Press

Quinn, J. & Dunham, A. (1983) On hypothesis testing in ecology and evolution. *The American Naturalist* **122**(5): 602–617.

Schrader-Frechette, K. & McCoy, E. (1994) What can ecology do for environmental management. *Journal of Environmental Management* **41**: 293–307.

Shima, J. (1999) Variability in relative importance of determinants of reef fish recruitment. *Ecology Letters* **2**: 304–310.

Simberloff, D. (1983) Competition theory, hypothesis-testing, and other community ecological buzzwords. *The American Naturalist* **122**(5): 626–635.

Simberloff, D. & Wilson, E. (1969) Experimental zoogeography of islands: the colonisation of empty islands. *Ecology* **50**(2): 278–296.

Underwood, A. (2000) Observations in ecology: you can't make progress on processes without understanding the patterns. *Journal of Experimental Marine Biology and Ecology* **250**: 97–115.

Vucetich, J., Peterson, R. & Schaefer, C. (2002) The effect of prey and predator densities on wolf predation. *Ecology* **83**(11): 3003–3012.

Willis, T. & Anderson, M. (2003) Structure of cryptic reef fish assemblages: relationships with habitat characteristics and predator density. *Marine Ecology - Progress Series* **257**: 209–221.

Wilson, S., Graham, N., Pratchett, M., Jones, G. & Polunin, N. (2006) Multiple disturbances and the global degradation of coral reefs: are reef fishes at risk or resilient? *Global Change Biology* **12**: 2220–2234.

M. AARON MACNEIL*
National Research Council, NOAA Panama City Laboratory, 3500 Delwood Beach Road, Panama City Beach, FL, 32408, USA

* Correspondence: Dr Aaron MacNeil Tel: +1 850 234 6541 ext. 257 Fax: +1 850 235 3559 e-mail: macneil@glau.ca

https://doi.org/10.1017/S0376892908004888 Published online by Cambridge University Press