Scientific method in environmental toxicology

Mark Crane and Michael C. Newman

Abstract: Environmental toxicologists, like all scientists, are presented with choices over the philosophical frameworks within which they work. However, most scientists do not receive formal training in scientific method and this may lead to inappropriate choices. The use of inductive and hypothetico-deductive approaches in science is described and the problems of each are discussed. Problems with induction include logical circularity and the selection of appropriate observations. Problems with hypothetico-deduction include generalizing from deductions, true predictions from false theories, the falsification of true theories, criteria for theory rejection, practical application in some sciences, and the potentially large number of unrefuted theories. Problems shared by both induction and hypothetico-deduction are the theory dependence of observations and the ahistorical nature of both of these explanations of scientific justification. The role of induction and hypothetico-deduction in environmental toxicology is discussed. Environmental toxicologists face two types of problems: (*i*) monitoring of the current fate and effect of a chemical, which is a historical and local problem that may often best be solved by an emphasis on hypothetico-deductive techniques; and (*ii*) prediction of the future fate and effects of chemicals, which usually has the objective of general applicability across time, species, and habitats, and may best be solved by an emphasis on inductive techniques. The systematic combination of both induction and hypothetico-deduction within a pluralistic framework is likely to yield the greatest progress in most areas of environmental toxicology.

Key words: scientific method, induction, hypothetico-deduction, monitoring, prediction.

Résumé : Les toxicologues environnementaux, comme tous les scientifiques, sont confrontés avec des choix sur les cadres philosophiques à l'intérieur desquels ils doivent travailler. Cependant, la plupart des scientifiques ne reçoivent pas de formation précise sur la méthode scientifique, et ceci peut conduire à des choix inappropriés. Les auteurs décrivent l'utilisation des méthodes inductive et hypothético-déductive en science et en discutent les problématiques. Les problèmes liés à l'induction incluent la circularité logique et la sélection des observations appropriées. Les problèmes liés à l'hypothético-déduction incluent la généralisation des déductions, les vraies prédictions issues de fausses théories ou de la falsification des vraies théories, les critères pour le rejet des théories, les applications pratiques dans certaines sciences et le grand nombre possible de théories jamais réfutées. Les problèmes communs aux deux méthodes, inductive et hypothético-déductive, sont la dépendance des observations par rapport aux théories et la nature historique de ces deux explications de la justification scientifique. Les auteurs discutent le rôle de l'induction et de l'hypothético-déduction en toxicologie environnementale. Les toxicologues environnementaux sont confrontés à deux types de problème : (*i*) le suivi (monitoring) du devenir actuel et effet de la substance chimique, un problème historique et local, qui peut souvent être résolu pour le mieux en mettant l'accent sur les techniques hypothético-déductives; et (*ii*) la prédiction du sort future et des effets chimiques, laquelle a généralement pour objectif l'applicabilité générale dans le temps, aux espèces et aux habitats, et peut être résolue au mieux en mettant l'accent sur les techniques inductives. La combinaison systématique des deux approches, inductive et hypothético-déductive, dans un cadre pluraliste est susceptible de conduire aux progrès les plus importants, dans la plupart des aspects de la toxicologie environnementale.

Mots clés : méthode scientifique, induction, hypothético-déduction, suivi, prédiction.

eceived January 10, 1996.
.ccepted April 18, 1996.

I. Crane.[1] Division of Biology, School of Biological ciences, Royal Holloway University of London, Egham, urrey, TW20 OEX, U.K.
I.C. Newman. University of Georgia, Savannah River cology Laboratory, P.O. Drawer E, Aiken, SC 29802, .S.A.

Author to whom all correspondence should be addressed.

Introduction

Until the middle of the 19th century, when the term scientist appears to have been coined, most scientists referred to themselves as natural philosophers (Whewell 1840). Now, although most modern philosophers respect science, many modern scientists have an exasperated contempt for philosophy (Medawar 1969) and may feel that it no longer encompasses science. There is a view that scientists at the beginning of their careers or in the prime of their professional life ought to concentrate only on science. Philosophy should be left either to philosophers or retired scientists who need a hobby.

Only the most naive professional scientist believes that his or her science is objective knowledge derived from measurements of reality. Yet if science is not as simple nor as unequivocal as this naive inductivism, then it is likely that the methods required to advance science are also neither simple nor unequivocal. In view of this, it is strange that the branch of philosophy relating to scientific method is rarely taught explicitly during the training of scientists. Environmental toxicologists will at some time in their careers learn the techniques, terminology, and ideas of ecology, toxicology, and chemistry. However, "...method as such [they] are presumably either expected to arrive with or to absorb spontaneously as they go along" (Gjertsen 1989). One associated danger is that practicing scientists may be unable to weigh the merits of different methodological approaches proposed by others.

There are persistent calls, mainly from academic quarters, for the adoption of a particular methodological approach in environmental toxicology. This recommended approach is usually the replacement of inductivism by hypothetico-deductivism (e.g., Cairns 1992; Forbes and Forbes 1994; Maltby and Calow 1989). These calls are made in an effort either to increase the efficiency of scientific discovery (Newman 1995) or improve the logic of scientific justification in environmental toxicology (Cairns 1992). Hypothetico-deductivism, it is argued, yields results more rapidly and efficiently, and provides a more logical framework for deciding on whether a scientific theory should be accepted.

Both hypothetico-deductivism and inductivism have strengths and weaknesses. It is important to recognize the strengths of each and use the most appropriate method for achieving the goals of environmental toxicology. The aim of this paper is to identify the relative merits of these different methodological approaches so that environmental toxicologists can choose methods that maximize both efficiency and logical soundness. A major theme of the paper is that neither induction nor hypothetico-deduction is a logically superior approach to the problems of environmental toxicology under all circumstances, but that in many situations inductive techniques offer more than some critics suggest.

What can we learn from the philosophy and history of science?

This section of the paper reviews the two main frameworks that philosophers of science have proposed to explain how scientists formulate and justify their theories. It is a simplification of the many views that have been advanced. We begin by addressing approaches based upon induction (the formulation of general theories from collections of individual observations) before moving on to approaches based upon hypothetico-deduction (the testing of a priori theories by comparing them with relevant data). Problems with each of these approaches are identified. The section ends by considering some sociological and psychological factors that must be taken into account if we are to understand how working scientists formulate and justify their theories in the real world.

For the sake of clarity, we should like to define science as widely as possible, as the attempt to contain the near infinite number and variety of phenomena in nature within a much smaller number of explanatory principles (Mayr 1982). Theory, model, and hypothesis are used synonymously, as there is no consensus on definitions for these terms in the scientific literature. We use all of them to describe an explanation of natural phenomena.

Classical induction

The public perception is often that science begins with observations, which the observer records as singular statements. For example, an environmental toxicologist might record from a single experiment that concentrations of cadmium greater than 0.1 µg/L adversely affect fathead minnow reproduction. Science also deals with universal statements about reality, such as "≥0.1 µg/L Cd adversely affects reproduction in all fish." How does one get from simple singular statements to more general universal statements?

We might be able to do so if we had many individual observations of the phenomenon under a range of different conditions and if none of the observations contradicted the universal statement. So, we might perform hundreds of toxicity tests with cadmium in different types of water, under different temperature and other physical regimes, with fish from different species and populations. If all of these tests showed that ≥0.1 µg/L Cd adversely affected reproduction in the test animals, then we might conclude that this concentration affects fish reproduction under all possible circumstances (although we shall show later why this conclusion is untenable).

This universal statement about the concentration of cadmium that affects fish reproduction is a theory from which we can predict previously unknown phenomena by deduction. Hence, if ≥0.1 µg/L Cd always affects fish reproduction, and brown trout is a previously untested fish species, then we can deduce that ≥0.1 µg/L Cd would also affect brown trout reproduction.

Nineteenth century inductivists added some sophistication to the original Aristotelian approach described above by expanding the concept of causation. According to this view, things happen because something causes them to happen, and it should be possible to develop rules for judging which of the many factors that occurred before an event was really its cause. Mill (1846) described four inductive experimental methods for determining which effects were caused by which antecedent events. The advantage of this approach is that it enables the use of a plurality of inductive methods. The approach has also been shown to work in practice. Epidemiology provides many case histories where Mill's rules have been used to great effect (Gjertsen 1989). Chapman (1995) gives an example of how a modern version of Mill's experimental methods might be used in an assessment of sediment pollution.

However, several fundamental problems with induction, outlined later in this paper, led many scientists and philosophers to propose that induction cannot be used to justify scientific explanations. Instead, the use of hypothetico-deductivism has been proposed as a superior alternative.

Hypothetico-deductivism
Most modern forms of hypothetico-deductivism follow a version of the falsificationist strategy proposed by Popper (1968). Popper correctly asserted that no theory can be conclusively proved but made the further assertion that a theory could be conclusively refuted. Popper's view was that science progresses through the falsification of theories and their replacement by superior ones. Falsificationists value a theory if it makes predictions that can be rigorously tested. If the predictions made by the theory are falsified, then the theory itself must surely be false.

One of the clearest expositions of this approach is by Platt (1964), in which the concept of strong inference was proposed. Strong inference consists of the following: (*i*) devising multiple alternative hypotheses; (*ii*) devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will exclude one of more of the hypotheses; (*iii*) carrying out the experiment so as to get a clean result; and (*iv*) recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain.

The hypothetico-deductive method as espoused by falsificationists requires theories to generate precise and clearly articulated predictions that are amenable to testing and falsification. Once a theory is falsified, it should be rejected and ad hoc modifications should not be made to overcome the reasons for falsification. In contrast with inductivists who wish to build up scientific knowledge piece by piece, falsificationists welcome bold conjectures that, if falsified, can be replaced in their entirety by other conjectures. Finally, scientists will arrive at a theory that withstands the most rigorous tests. This theory is not necessarily true, but it is better than the ones that preceded it. Science progresses by developing theories that explain more and are hence more falsifiable. Falsificationists have no strongly held opinions on how one might arrive at a conjecture: it could be as a result of inspiration, perspiration, or any other process, including induction (Loehle 1990, 1994; Peters 1991). They are preoccupied with how a scientific theory can be justified and are less interested in how it is discovered. This can be contrasted with inductivism, which seeks to explain both scientific discovery and justification. What matters to falsificationists is that once the theory has been formulated, it should be amenable to falsification.

Despite this emphasis on falsification, confirmation also plays an important role in the falsificationist's view of science (Chalmers 1982). Significant scientific progress is made when highly speculative theories are confirmed or when cautious theories that were considered unproblematic are falsified. The falsification of highly speculative theories or the confirmation of cautious theories does not lead to as much progress.

Hence, we may formulate the simple hypothesis that ≥0.1 μg/L Cd adversely affects the reproduction of all fish species. According to falsificationists, we should seek to falsify this hypothesis by making risky predictions. For example, we may choose to test a species that we know is very tolerant to other heavy metals or we might choose exposure conditions that minimize the bioavailability of cadmium, such as hard water or the presence of sediment. If any of our tests show that even a single fish can reproduce normally if exposed to ≥0.1 μg/L Cd, then the hypothesis should be rejected.

Problems with induction and hypothetico-deduction
Neither induction nor hypothetico-deduction is a problem-free approach. Problems with induction include the circularity of arguments for its justification and the selection of a sufficiently large number and wide range of observations. Problems with hypothetico-deduction include the inadequacy of deduction alone for most scientific generalizations, the possibility that false theories could lead to true predictions and that true theories could be falsified, the extent to which a theory needs to be falsified before it is rejected, the difficulty of practically applying hypothetico-deduction to some sciences, and selection from a potentially infinite number of unrefuted theories. Problems shared by both induction and hypothetico-deduction are the theory dependence of observations and the ahistorical nature of both of these explanations of scientific justification. Each of these problems is addressed in more detail below.

Problems with induction
A fundamental flaw in the inductive approach is its logical circularity. We may inductively state that the reproduction of all fish species will be adversely affected when they are exposed to ≥0.1 μg/L Cd. Such a statement asserts that future instances of which we have had no experience must resemble past instances of which we have had experience and that the course of nature will always remain the same (Hume 1748). This is a circular argument that relies upon induction to justify induction. Natural phenomena can and do change, and so past experiences cannot be relied upon to predict future phenomena unerringly. For example, we can never be certain that the reproduction of all fish will always be affected at ≥0.1 μg/L Cd, even if this has always happened in the past. Tomorrow we may discover a population that has developed a particularly high tolerance to heavy metals. Next year we may find a new species with a physiology that is not susceptible to cadmium. This difficulty with the inductivist approach has traditionally been called the "problem of induction."

There are several other problems with induction. For example, what constitutes a sufficiently large number of observations before universal statements can be made? Perhaps some phenomena require fewer observations to establish knowledge (Chalmers 1982). Presumably only a few large oil spills are required before scientists and the public recognize the short-term danger of such incidents to aquatic ecosystems. We might compare this with the larger number of observations required to establish whether the discharge of oil-contaminated ballast at sea causes pollution.

In addition to the number of observations required, we might ask how wide the range of different conditions should be under which we make these observations. When looking at cadmium toxicity to fish, do we need to make observations at different water hardness values, on different populations of the same species, and at different stages of the lunar cycle? The answers are probably "yes" to the first, "possibly" to the second, and almost certainly "no" to the third. But how do we select the important variables?

These and other problems with the use of induction as a means to knowledge have led some philosophers and scientists to deny that induction can have a role in justifying scientific theories (Popper 1968). Hypothetico-deduction is claimed to be a superior approach. However, despite the widespread acceptance of hypothetico-deductivism as a philosophical framework by environmental toxicologists and ecologists (e.g., Cairns 1992; Forbes and Forbes 1994; Loehle 1987; Maltby and Calow 1989; Newman 1995; Peters 1991; Salt 1983; Simberloff 1983; Underwood 1991; Wiens 1984; but see review of debate on "Popperphilia" in community ecology in McIntosh 1987), this approach also has several serious problems.

Problems with hypothetico-deductivism

Deduction alone cannot provide an adequate model for all scientific reasoning. This is because a purely deductive argument depends upon the provision of the answer in the initial premisses. Most sciences, including environmental toxicology, attempt to produce universal statements that apply to unmeasured members of the true population and to future, as well as current and past, events. It is impossible for a purely deductive conclusion to be drawn from premisses that include statements about the future or about measurements upon only a sample of the population. Because of this, induction is usually involved in at least a supporting role in the hypothetico-deductive approach, especially when inferential statistics are involved.

There is also no logical reason why false theories should not lead to novel and true predictions, as several historical examples demonstrate (Gjertsen 1989). The converse of this is that true theories can be falsified by false observations or perhaps by a lack of statistical power (Mentis 1988). Falsificationism has some difficulty with assessing the probabilistic hypotheses that are most common in many biological disciplines, including environmental toxicology (Howson and Urbach 1993; Toft and Shea 1983). Deterministic hypotheses about the speed of light, the density of water, or the molecular structure of DNA make quite precise predictions that can be tested and either confirmed or refuted. Any small differences between measurement and prediction are usually attributed to experimental error caused by imperfect measuring devices. However, hypothesis testing of phenomena with biological individuals, populations, and communities is generally not of this type. Instead, it consists of statistical hypothesis testing. Only probabilistic confirmations or refutations are possible with this approach: a hypothesis is rejected if it is relatively unlikely compared with other outcomes of the experiment. Since low power is a feature of many experiments in ecology and environmental toxicology (Forbes and Forbes 1994; Suter 1993; Toft and Shea 1983), the chance of committing a type II error (accepting the null hypothesis when it is false) is often high.

Another problem with the emphasis on falsification in modern hypothetico-deductivism arises when we ask how much of a theory needs to be falsified. Most scientific theories are complex structures that depend upon a collection of universal statements, initial conditions, and auxiliary assumptions (Chalmers 1982). If a prediction made by a theory is false, it may be due to an error in an auxiliary assumption rather than in the main body of the theory itself. Because of this, scientists are usually able to deflect a falsification onto a less important aspect of their theoretical construct and hence protect the main theory (Lakatos 1974).

Some seemingly competitive theories might also be complementary if, in fact, they each relate to a particular situation (Quinn and Dunham 1983; Mentis 1988). For example, many chemical partitioning theories depend upon system equilibrium as an auxiliary assumption (e.g., DiToro et al. 1991). If a prediction of toxicity based on such a theory was falsified in a sediment toxicity test, does it mean that the theory is wrong, the auxiliary assumption of equilibrium was not met, or some other assumption about initial or auxiliary conditions was false?

It is also possible that the less deterministic sciences are less amenable than others to a hypothetico-deductive approach. For example, it has been argued that ecology, which shares many of the aims and tools of environmental toxicology, may be served poorly by a rigid adherence to hypothetico-deductivism. This is because hypotheses in ecology tend not to be lawlike (in the sense of laws in physics), but are often simply principles (e.g., the principle of competitive exclusion). This can make it very difficult to specify practically the range of phenomena that the theory represents, a procedure known as operationalization (Peters 1991). It is also sometimes difficult to give equal weight to alternative hypotheses, as suggested by Platt (1964), because some hypotheses may be better developed, easier to test, or more value laden than others (Roughgarden 1983). This makes the status of alleged theory confirmations in environmental toxicology and ecology problematic (Shrader-Frechette and McCoy 1993).

One final objection to the dogmatic prescription of hypothetico-deductivism is that it simply does not provide a scientist with any tools for choosing between unrefuted theories. Howson and Urbach (1993) argue that refuting a false theory is of little help, because it leaves an infinite class of theories that have not been refuted. Which one of these can reasonably be considered as true? Bayesians, such as Howson and Urbach, believe that scientists already have an intuitive and inductive system for ranking competing hypotheses in some sort of value order. Scientists do not consider all possible competing hypotheses to be of equal worth. Bayesians believe that scientists are more interested in questions about the probable plausibility of a hypothesis, rather than the probable frequency of particular outcomes. Bayesian analysis asks scientists to accept that probability depends upon the confidence that they are prepared to invest in a proposition (Jefferys and Berger 1992). In other words, scientific reasoning is both subjective and probabilistic. This is a difficult pill for most scientists to swallow after a professional training that, correctly or incorrectly, emphasizes the removal of subjective assessment from the scientific process. As a result of this, the use of subjective Bayesian techniques has remained controversial, although there appears to be a very wide range of problems in environmental toxicology that might be amenable to this approach (e.g., Reckhow and Chapra 1983).

Problems shared by induction and hypothetico-deduction

A problem shared by both induction and hypothetico-deduction is the theory dependence of observation (Fagerström 1987). Observations cannot be made independently from theory: we resort to theory before we are able to select appropriate observations. If this is the case, then theory precedes observation and not vice versa as inductivists suggest. Even the observation statements that we make presuppose theory. If we state that cadmium adversely affects fish reproduction then we are assuming, amongst other things, that there is a group of organisms that can be classified as fish (a taxonomic theory). If we move on to only slightly more complex observation statements such as "increasing pH was found to increase cadmium toxicity to fish," then it should become clear that considerable chemical theory about the unobservable behaviour of hydrogen ions is involved.

If we accept that observations are theory dependent, then it is clear that observations can only be as reliable as their associated theories. If the theory is flawed, then the observation may also be false. Supporters of hypothetico-deductivism also have to admit the possibility that even a seemingly conclusive refutation may be incorrect, because the theory on which it is based is flawed. This problem cannot be overcome by making observations more objective in some sort of falsificationist manner (Underwood 1991). Underwood (1991) claims that one can be sure that observations reflect the real world by developing a null hypothesis for one's observations and attempting to falsify it. Unfortunately, this procedure merely moves all of the problems previously identified with hypothetico-deductivism back one stage in the procedure. There really does not seem to be any way of establishing the objective reality of observations. Popper (1968) was aware of this problem:

> "The empirical basis of science has thus nothing 'absolute' about it. Science does not rest upon solid bedrock. The bold theories of its structures rise, as it were above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being."

Social and psychological factors

Perhaps the most telling criticism of dogmatic calls to adopt a particular method is simply that this is not the way that successful scientists seem to go about their work. Scientific discovery and justification may be better understood by studying the history of individual sciences, rather than just their logic (Feyerabend 1970; Mayr 1982). This is not to suggest that a method is right just because scientists have tended historically to follow it. However, it is as pointless to recommend standards for scientists that are psychologically and historically unrealistic, as it is to yield to extreme relativism and the "anything goes" school of thought (Bechtel and Richardson 1993).

Do scientists follow a method?

Both modern inductivists and hypothetico-deductivists can provide us with reasonably coherent accounts of how scientific explanations might be discovered and (or) justified. But is this really the way that scientists go about the job of science?

Fagerström (1987) suggests that psychological reasons for retaining a theory, such as its apparent consistency, productivity, simplicity, beauty, and elegance, may far outweigh considerations of its truth, and that even great scientists will develop and promote a theory despite evidence to the contrary.

The positive side of this psychological phenomenon is that theories may require time to mature before they can produce testable predictions (Kuhn 1962; Loehle 1987, 1988). For example, Newman (1996) argues that some areas of environmental toxicology are currently in a phase in which an insistence on rigorous falsification would lead to their premature rejection, while other areas should have passed from this phase and now be exposed to more stringent tests. On the negative side, scientists, like all humans, are prone to confirmation bias: the psychological desire to confirm their theories rather than attempt to falsify them. This can lead to a loss of normal levels of scientific objectivity and to what Rousseau (1992) has termed pathological science. Hence, adherence to an immature theory may represent either a process of positive nurturing until the theory is precise and predictive, or a stubborn adherence to a hopelessly vague and qualitative concept on the part of the scientist.

Normal and revolutionary science

Kuhn (1962) provided the first coherent account of the sociological setting within which most scientists work and how this leads to observed historical patterns of scientific method. According to Kuhn, most scientists solve puzzles within a well-defined "paradigm," consisting of the theoretical assumptions and the techniques for their application that are adopted by a scientific community. Normal science is what scientists do within a paradigm to articulate and extend it by detailed mathematical, field, or laboratory work. This articulation will often include the collection of confirmatory data that can be explained by the paradigm (Murray 1986).

However, at times paradigms break down because they are unable to provide answers to critical problems or because anomalies arise. Puzzles that strike at the heart of the paradigm remain unresponsive to all efforts to solve them. At such a time there is a period of crisis within the paradigm. Creative scientists will then seek ways of overthrowing the prevailing paradigm and replacing it with another that can answer the critical problems and remove the anomalies, if such a new paradigm is available. This is a period of scientific revolution, which is then followed by a continuation of normal science within the newly adopted paradigm.

Scientific research programmes

The history of science, as described by many philosophers and historians of science since Kuhn, suggests that theories are neither discovered nor justified in the way that either inductivists or hypothetico-deductivists suggest. Instead, theories appear to evolve into complex organized structures over time. Lakatos (1974) described these organized structures as "scientific research programmes," containing several identifiable features. The "negative heuristic" describes a core of basic assumptions that cannot be rejected, equivalent to Kuhn's (1962) concept of a paradigm.

A protective belt of auxiliary assumptions and initial conditions surrounds this core. The "positive heuristic" of a research programme identifies ways in which the core can be used to predict novel phenomena and solve problems. It consists of a partially articulated set of suggestions or hints on how to change and develop the refutable variants of the research programme and to modify the protective belt to make it more sophisticated.

In a good scientific programme, any modification is allowable except an ad hoc one. In other words, all modifications made to a programme must be testable and not merely contrived to overcome problems with the programme. The other type of modification that Lakatos rules out is one to the core of the programme. Lakatos (1974) allows scientists more flexibility than Popper when developing their research programmes. In an approach to entire programmes similar to Loehle's (1987) approach to theory, Lakatos suggests that immature but growing research programmes should at least initially be sheltered from the full force of falsification.

Appropriate methodological frameworks for environmental toxicologists

The preceding review of inductive and hypothetico-deductive frameworks, and the psychological and sociological factors that influence scientists is as relevant to environmental toxicology as it is to the fundamental sciences. In this section we identify the two main goals of environmental toxicology and suggest that different methodological emphases may be most appropriate for achieving each of these goals.

A complete historical account of environmental toxicology has not yet been written, although some useful contributions towards it have been made (e.g., Halffman 1995; Mount 1995). Despite this, it is possible to identify at least some of the paradigms that have developed within the field. The core paradigms of environmental toxicology consist, amongst others, of theory from evolutionary biology, genetics, molecular biology, chemistry, and human toxicology. An environmental toxicologist who does not accept the bulk of these core ideas would be regarded as eccentric or ill-informed by other members of the scientific community. The protective belt that surrounds this core consists of assumptions and opinions about such things as appropriate statistical and risk assessment methods, thresholds, and the potential contribution of ecological concepts. Environmental toxicologists can hold quite different views on these issues but still be regarded as belonging to the same scientific community. In fact, the debate over such issues forms much of the positive heuristic of the science, as its practitioners strive to achieve agreed goals.

What are these goals of environmental toxicology? The scientific aim is to build knowledge about the fate and effects of contaminants in the biosphere, based on explanatory principles. Two practical objectives that support this aim are (*i*) the measurement of the current fate and effect of chemicals in particular places, and (*ii*) the general prediction of the possible fate and effects of chemicals in the future. In other words, the problem faced by environmental toxicologists is either one of retrospective assessment or of predictive assessment (Maltby and Calow 1989; Suter 1993).

The determination of the current fate and effect of single chemicals or mixtures (a retrospective impact study) is usually a historical and spatially well-defined problem that may best be solved with an emphasis on the hypothetico-deductive approach. The prediction of the fate and effects of chemicals usually has the objective of universal or at least very wide applicability across time, species, and habitats, and may best be solved by emphasizing inductive approaches. These approaches are addressed in more detail below.

Retrospective impact studies

In this section we describe the nature of impact studies and some advantages of using a hypothetico-deductive approach within them. The influence of ad hoc theorizing is then discussed. Finally, conditions are identified for which a greater emphasis on induction may be of most use.

The nature of impact studies

An impact study in environmental toxicology is usually both local and historical in nature. The investigator is interested in chemical fate and effects at specific and unique sites, rather than at generalized or universal localities. He or she is also interested in the historical fate of contaminating chemicals and in the history of potentially affected organisms. Impact studies may be triggered by observations on sources (e.g., an oil spill), effects (e.g., a fish kill), or exposure (e.g., the detection of pesticides in surface waters). This multiplicity of reasons for conducting an impact study contrasts with the single source-driven rationale for predictive studies.

In the predictive studies described later, the investigator knows the source (e.g., a new chemical or effluent) but needs to estimate the level of exposure experienced by organisms and the toxic effects that result from this exposure (Suter 1993). This can be contrasted with an impact study on a known source, in which the task of the investigator is also to discover the level of exposure and effects, but usually by direct measurement rather than estimation. With known effects in an impact study, the task is to discover the type of exposure and source that caused the effects; with known exposure, the task is to identify both the source of the exposure and whether the exposure causes any effects (Suter 1993). Hence, the explanation of cause and effect is an important aim of impact studies, whatever their genesis.

The notion of causality must be treated with some caution. The criteria for causation have been established for nearly 250 years (Hume 1748): cause and effect should occur together in space and time, the effect should follow the cause, and an effect should always occur when the cause is present. Most of us use commonsense notions of cause and effect in our everyday lives, but there is considerable evidence to show that many of these commonsense notions fail to describe the real world (Nisbett and Ross 1980). Possible reasons for this are that (*i*) people rely on the criterion of resemblance (big effects must have big causes); (*ii*) new causal models are built by uncritical analogy with existing ones; (*iii*) plausible causal schemes, once identified, are maintained in the face of contrary evidence; (*iv*) there is an assumption that the information at hand is sufficient even when it is not; (*v*) the most apparent of prospective causes is usually selected from an available list; and (*vi*) people are too willing to look for single causes.

Hence, environmental toxicologists must exercise greater than normal vigilance when seeking to ascribe particular causes to specific effects and not rely upon common sense alone. In fact, it has been argued that environmental scientists should abandon the concept of cause altogether and replace it with the more restricted, but operational, search for predictable regularities (Peters 1991). The complexity of even this more restricted goal requires both imagination and a systematic approach. A system that has proved effective in the past is to adopt the method of multiple working hypotheses, rather than favour one theory above others (Chamberlin 1897), and then to follow the principles of strong inference (Platt 1964).

Advantages of the hypothetico-deductive approach in impact studies

Strong inference is probably the most rigorous form of hypothetico-deductivism. If followed consistently, then it is the method most likely to identify the strongest candidate from a list of alternative explanatory hypotheses. It is particularly useful when several explanatory hypotheses can be tested simultaneously or sequentially, and when each of these hypotheses makes quantifiable and falsifiable predictions that differ considerably from one another.

An example of this might be an analysis of factors responsible for the reduced abundance of amphipods in an acidic lake (an example of an effects-driven study). A range of plausible alternative hypotheses is available. Possible causal mechanisms are the direct effects on the animals of reduced pH, the liberation of heavy metals from soils in the catchment by acid rain, changes in fish predatory behaviour due to reduced zooplankton abundance, reductions in algal food material, natural stochastic variation in amphipod abundance, and so on. Each of these hypotheses should be formulated so that precise mutually exclusive predictions are made that are amenable to falsification. A logical tree should then be drawn up so that alternatives and crucial experiments can be identified. The experiments should be performed, and then subhypotheses should be formulated and tested, until eventually all but one of the hypotheses is rejected. The remaining hypothesis, which may be some sort of composite of previous hypotheses due to interactions between different causal variables, should then be subjected to the most rigorous tests of its predictions. It should never be accepted as the truth but only as the best approximation to truth achieved so far.

Ad hoc theories

The theoretical constructs within a historical analysis need not necessarily be predictive: strong inference is not the only possible approach (Mayr 1982; Peters 1991). However, there is a tendency with historical analyses that do not follow the procedures of strong inference to construct a likely story that provides us with only one out of a range of logical possibilities (Loehle 1988; Murray 1986; Peters 1990). There is then a danger of precipitate explanation (Newman 1995). If classical statistical approaches are used for analysis then it is important that the hypothesis is constructed before data are collected rather than invented in some ad hoc manner after measurements have been gathered. Prior hypothesizing allows the experimenter to define the size of the biological differences that he or she wishes to detect and to design a study with sufficient power to detect such effects. There is nothing wrong with using a posteriori hypotheses generated from statistical analyses of data to design further studies. However, hypotheses developed in such a way are not confirmed by the data used in their derivation. They should be used only for the generation of further predictions for testing. If they are not used in this way, they "...become mere ad hoc hypothesising for individual data sets" (Mentis 1988).

However, one can imagine that some ad hoc hypotheses are more plausible than others, and some quite plausible ones may no longer be testable because they were temporally dependent. Must we then abandon even plausible modifications to our original hypothesis? Bayesians take a different view towards ad hoc theorizing and believe that the validity of an ad hoc theory should be related to its quantifiable probability (Howson and Urbach 1993).

The role of induction in impact studies

There are several potentially productive interactions between induction and hypothetico-deduction in impact studies (Mentis 1988). These include the collection of data and their analysis by regression techniques prior to hypothesis formulation; the screening of multiple causes and effects to produce a short list for experimental manipulation; the comparison of many competing theories to provide an objective short list that can be assessed by hypothetico-deductivism; and the development of predictive models for subsequent hypothesis testing, a process that may involve an iteration between inductive modelling and hypothesis testing until the required precision of prediction is attained.

It appears that humans have a subconscious philosophical preference for discrete classifications, a characteristic that scientists, as humans, share (Loehle 1988). The need to calculate a no observed effect concentration (NOEC) in bioassays probably reflects a desire to classify particular chemical concentrations or environmental samples as good or bad, rather than better or worse. The desire to classify aggregations of organisms into spatially and typologically discrete community types may reflect a similar psychological bias. Environmental managers also often see a need to produce simple classifications so that they can provide understandable explanations for politicians and members of the public untrained in science. Hypothesis testing using the rules of strong inference should normally be a powerful approach to impact studies where sensible discrete boundaries can be drawn, with or without the support of inductive observations.

However, many phenomena of interest to environmental toxicologists should not be placed in discrete categories but at some point along a continuum. The mixing zones near effluent outfalls and oil contamination around drilling installations are just two examples. Such questions are probably best answered by inductive modelling. Hypothesis testing in these cases simply does not provide efficient answers (Suter 1993). There is less information in the statement that chemical concentration x at site A causes significantly greater effects than chemical concentration y at site B, than in a statement about the function of the response y on x across all contaminated sites.

Prediction of future fate and effects of chemicals

A suitable methodological framework for predicting the future fate and effects of chemicals must also involve a large element of induction, if only because many of the technical and conceptual tools required for mechanistic solutions are simply unavailable. They may never become available. Until they do, a statistical black box, based upon linear modelling or more sophisticated multivariate techniques, is one of the few options available to us. In this section we try to define what environmental toxicologists mean when they refer to prediction. The advantages and disadvantages of induction for meeting the goals of prediction are then identified.

The nature of predictive studies in environmental toxicology

Mayr (1982) identifies an important difference between logical and temporal prediction. Logical prediction is the conformance of individual observations with a theory, according to the hypothetico-deductive approach. However, when environmental toxicologists talk about prediction, they usually mean temporal prediction or an inference from the present to the future. They particularly want to know how a new chemical will behave in different habitats and how it will affect a wide range of different organisms. It may be that precise temporal predictions are rarely possible in the biological sciences owing to the combined effects of randomness, uniqueness, complexity, emergent properties, and the magnitude of stochastic perturbations (Mayr 1982). Unfortunately, the selection of only soluble problems (Medawar 1969; Loehle 1990) is not a luxury offered to environmental toxicologists working outside academic departments. Those employed in industry and government, as technical or practical ecotoxicologists (Newman 1996), simply must try to predict the effects of novel compounds on biological populations and ecosystems. There is no other discipline to claim their "leftovers" (Slobodkin 1988).

Advantages of induction for prediction

Predictive environmental toxicology needs to identify system attributes for prediction that are operational and preferably simple and quantitative. Concepts that are currently nonoperational, such as ecological stability, stress and health, should not be selected until they are operationalized by their proponents (Crane 1995; Peters 1991; Shrader-Frechette and McCoy 1990, 1993; Underwood, 1986). Once operationalized, empirical modelling based on regression and multivariate approaches can be used to relate dependent variables to possible independent variables.

There are strong parallels between the instrumentalist approach recommended here and the predictive ecology championed by Peters (1986, 1988, 1990, 1991). Mechanistic or holistic explanations suggested by empirical relationships between variables should be welcomed but are not a necessary objective. Indeed, probably the most important strength of induction for environmental toxicologists is the ability to ignore cause and effect. This may be necessary for pragmatic reasons alone if a rapid prediction is required, but underlying processes have not been fully discovered.

More than 90% of the current predictive studies in environmental toxicology rely on single-species test systems in which the death of fish or invertebrates is measured (Maltby and Calow 1989). There have been many criticisms of simple single-species testing as a predictive tool for environmental effects. These criticisms have been voiced by both holists and mechanists (Cairns 1986, 1992; Cairns and Niederlehner 1987; Cairns et al. 1992; Forbes and Forbes 1994; Giesy and Odum 1980). However, the usual holistic solution proposed to overcome these perceived difficulties, i.e., the use of multispecies test systems, is not a logical one. The criticism that results from single-species tests should not be generalized across populations, species, and habitats is a complaint against the use of induction. However, the use of multispecies test systems does not help environmental toxicologists to escape induction. It simply gives them a more complex and costly set of results for generalization across populations, species, and habitats other than those found in the multispecies system (Maltby and Calow 1989).

On the other hand, claims that mechanistic analyses at the level of the organism may be used to predict effects at the population and community level founder upon the logical and practical incommensurability of functional phenomena such as organisms and typological phenomena such as populations and communities (Loehle 1988; Shrader-Frechette and McCoy 1990). Mechanistic explanations also tend to concentrate on material and efficient causes and ignore the possible influence of formal and final causes in positive ecosystem feedbacks (Bechtel and Richardson 1993; Mayr 1982; Suter 1995; Ulanowicz 1988; 1990). In addition, mechanistic analyses often depend upon a restricted number of explanatory causes, when multiple causation is the ecological norm (Peters 1991). These problems with the mechanistic approach to environmental toxicology could lead to never-ending research (Peters 1991), the fondling of research problems (Gjertsen 1989), and a failure to answer pressing questions about the environmental effects of chemicals. Mechanistic strategies do have an important role in environmental toxicology, but probably only when systems can be decomposed into units that are governed mainly by intrinsic factors, and the phenomenon of interest can be localized within one or more of these specific units (Bechtel and Richardson 1993).

Under such circumstances, induction through empirically derived statistical relationships probably offers the best hope for general predictions with an applied use. This approach, in which data from the simplest practicable level are used to predict effects at the level of interest, has been termed "empirical holism" (Peters 1991). The independent variables used for prediction should be the simplest that provide a sufficiently precise prediction. Recent advances in the use of quantitative structure–activity relationships

(QSARs) and other empirical extrapolation techniques suggest that in environmental toxicology single-species tests are sufficient for this purpose in most cases (e.g., Bradbury 1994; van Leeuwen et al. 1992; Wagner and Løkke 1991). Opponents of this view (e.g., Forbes and Forbes 1994) need to provide evidence that more complex approaches can offer more precise predictions.

Disadvantages of induction for prediction

As well as sharing the strengths of predictive ecology, the inductive approach to predictive environmental toxicology shares some of its weaknesses. First, there will always be some residual error in probabilistic predictions, leading to some degree of imprecision; one of the main aims of instrumentalist research must therefore be to minimize this residual error. Second, a correlation does not logically entail causation, although it often suggests it. It is likely that certain phenomena simply cannot be predicted with any precision due to hierarchical effects on causality (Ulanowicz 1988). This should be recognized and phenomena should be selected for prediction that are most amenable to one-to-one mapping (Suter 1995). In practice, this may mean an ability to predict the effects of contaminants on phenomena such as total biomass and the abundance of individual populations of selected organisms, but not the effects of contaminants on phenomena such as community structure. It also requires the selection of measurements at temporal scales that reflect those of the phenomenon of interest: i.e., the scales of possible cause and effect should be commensurate (Harris 1994).

Conclusions

Scientific method is how we go about doing science. Some philosophers and scientists have suggested that there is only one truly scientific method (e.g., Popper 1968, 1969). Others suggest that there are no logically defensible methods at all and that anything goes (Feyerabend 1975). However, if anything goes, then everything stays, including the obviously bad (Mentis 1988). Therefore, we need at least some rules for determining what is and what is not acceptable. An acceptable approach in an applied science like environmental toxicology is simply one that demonstrably works and achieves agreed goals. The two main methodological frameworks proposed for science, induction and hypothetico-deduction, both have advantages and disadvantages. However, neither one can provide a complete answer to the problems faced by scientists and both have disadvantages, especially when used in isolation.

Problems with induction include the circularity of arguments for its justification and the selection of a sufficiently large number and wide range of observations. Problems with hypothetico-deduction include the inadequacy of deduction alone for most scientific generalizations, the possibility that false theories could lead to true predictions and true theories could be falsified, the extent to which a theory needs to be falsified before it is rejected, the difficulty of practically applying hypothetico-deduction to some sciences, and problems with selecting from a potentially infinite number of unrefuted theories. Problems shared by both induction and hypothetico-deduction are the theory dependence of observations and the ahistorical nature of both of these explanations of scientific justification.

An exclusive distinction between inductive and deductive approaches in biology is clearly artificial (Calow 1987). In environmental toxicology, as in ecology, a pluralistic approach is likely to yield the greatest progress (McIntosh 1987), so long as this pluralism is embedded in a systematic structure. Induction without deduction is simply a collection of facts, while hypothetico-deduction cannot provide scientific generalizations without at least some element of induction, a fact that even Popper came to recognize (Losee 1993). Historical and sociological analyses of scientific progress also support the view that inductive and hypothetico-deductive approaches are complementary. The question is one of emphasis rather than exclusion.

The *justification* of theory in environmental toxicology should clearly benefit from a structured and logical approach. The hypothetico-deductive method, as exemplified by strong inference, can be a rigorous tool in retrospective impact studies, so long as its potential weaknesses are not ignored. However, a focus on hypothetico-deductivism is probably neither a logical nor a practical approach to most predictive studies. An emphasis on inductive techniques seems to promise the most success in this area, although sole use of classical statistical techniques may overlook useful contributions from Bayesian theory. The *discovery* of theory in environmental toxicology should also benefit from a structured approach, but it is unlikely that hypothetico-deduction can form the basis of this.

Perhaps the most pressing need in evaluating the merits of different methods in environmental toxicology is a rigorous historical analysis that describes its ontogeny, its agreed goals, and its identifiable successes. As Chalmers (1986) states, "There is no universal method. There are no universal standards. But there are historically successful standards implicit in successful practices."

Acknowledgements

The contribution to this manuscript by M. Newman was supported by Contract DE-AC09-76SR00-819 between the U.S. Department of Energy and the University of Georgia (Savannah River Ecology Laboratory).

References

Bechtel, W., and Richardson, R.C. 1993. Discovering complexity. Decomposition and localization as strategies in scientific research. Princeton University Press, Princeton, N.J.

Bradbury, S.P. 1994. Predicting modes of toxic action from chemical structure: an overview. SAR QSAR Environ. Res. **2**: 89–104.

Cairns, J., Jr. 1986. The myth of the most sensitive species. Bioscience, **36**: 670–672.

Cairns, J., Jr. 1992. Paradigms flossed: the coming of age in environmental toxicology. Environ. Toxicol. Chem. **11**: 285–287.

Cairns, J., Jr., and Niederlehner, B.R. 1987. Problems associated with selecting the most sensitive species for toxicity testing. Hydrobiologia, **16**: 101–112.

Cairns, J., Jr., McCormick, P.V., and Niederlehner, B.R. 1992.

Estimating ecotoxicological risk and impact using indigenous aquatic microbial communities. Hydrobiologia, **237**: 131–145.
Calow, P. 1987. Towards a definition of functional ecology. Funct. Ecol. **1**: 57–61.
Chalmers, A.F. 1982. What is this thing called science? Open University Press, Milton Keynes, U.K.
Chamberlin, T.C. 1897. The method of multiple working hypotheses. J. Geol. **5**: 837–848.
Chapman, P.M. 1995. Extrapolating laboratory toxicity results to the field. Environ. Toxicol. Chem. **14**: 927–930.
Crane, M. 1995. Is there a place for ecology in ecotoxicology? SETAC News, **15**(2): 19–20.
DiToro, D.M., Zarba, C.S., Hansen, D.H., Berry, W.J., Swartz, R.C., Cowan, C.E., Pavlou, S.P., Allen, H.E., Thomas, N.A., and Paquin, P.R. 1991. Technical basis for establishing sediment quality criteria for nonionic organic chemicals using equilibrium partitioning. Environ. Toxicol. Chem. **10**: 1541–1583.
Fagerström, T. 1987. On theory, data and mathematics in ecology. Oikos, **50**: 258–261.
Feyerabend, P. 1970. Philosophy of science: a subject with a great past. *In* Historical and philosophical perspectives of science. *Edited by* R. Stuewer. University of Minnesota Press, Minneapolis, Minn. pp. 172–183.
Feyerabend, P. 1975. Against method. Humanities Press, London, U.K.
Forbes, V.E., and Forbes, T.L. 1994. Ecotoxicology in theory and practice. Chapman and Hall, London, U.K.
Giesy, J.P., Jr., and Odum, E.P. 1980. Microcosmology: introductory comments. *In* Microcosms in ecological research. *Edited by* J.P. Giesy Jr. United States Department of Energy, Washington, D.C. Symp. Ser. 52 (Conf-781101). pp. 1–13.
Gjertsen, D. 1989. Science and philosophy, past and present. Penguin Books, London, U.K.
Halffman, W. 1995. The boundary between ecology and toxicology: a sociologist's perspective. *In* Ecological toxicity testing: scale, complexity and relevance. *Edited by* J. Cairns Jr. and B.R. Niederlehner. Lewis Publishers, Boca Raton, Fla. pp. 11–34.
Harris, G.P. 1994. Pattern, process and prediction in aquatic ecology. A limnological view of some general ecological problems. Freshwater Biol. **32**: 143–160.
Howson, C., and Urbach, P. 1993. Scientific reasoning. The Bayesian approach. 2nd ed. Open Court Press, Chicago, Ill.
Hume, D. 1748. An enquiry concerning human understanding. Open Court Publishing Co., Chicago, Ill.
Jefferys, W.H., and Berger, J.O. 1992. Ockham's razor and Bayesian analysis. Am. Sci. **80**: 64–72.
Kuhn, T.S. 1962. The structure of scientific revolutions. University of Chicago Press, Chicago, Ill.
Lakatos, I. 1974. Falsification and the methodology of scientific research programmes. *In* Criticism and the growth of knowledge. *Edited by* I. Lakatos and A. Musgrave. Cambridge University Press, Cambridge, U.K. pp.
Loehle, C. 1987. Hypothesis testing in ecology: psychological aspects and the importance of theory maturation. Q. Rev. Biol. **62**: 397–409.
Loehle, C. 1988. Philosophical tools: potential contributions to ecology. Oikos, **51**, 97–104.
Loehle, C. 1990. A guide to increased creativity in research — inspiration or perspiration? Bioscience, **40**: 123–129.
Loehle, C. 1994. Discovery as a process. J. Creative Behav. **28**: 239–250.
Losee, J. 1993. A historical introduction to the philosophy of science. 3rd ed. Oxford University Press, Oxford, U.K.
Maltby, L., and Calow, P. 1989. The application of bioassays in the resolution of environmental problems; past, present and future. Hydrobiologia, **188/189**: 65–76.
Mayr, E. 1982. The growth of biological thought: diversity, evolution and inheritance. The Belknap Press of Harvard University Press, Cambridge, Mass.
McIntosh, R.P. 1987. Pluralism in ecology. Annu. Rev. Ecol. Syst. **18**: 321–341.
Medawar, P. 1969. The art of the soluble. Harmondsworth, London, U.K.
Mentis, M.T. 1988. Hypothetico–deductive and inductive approaches in ecology. Funct. Ecol. **2**: 5–14.
Mill, J.S. 1846. System of logic. London, U.K.
Mount, D.I. 1995. Development and current use of single species aquatic toxicity tests. *In* Ecological toxicity testing: scale, complexity and relevance. *Edited by* J. Cairns Jr. and B.R. Niederlehner. Lewis Publishers, Boca Raton, Fla. pp. 97–104.
Murray, B.G., Jr. 1986. The structure of theory, and the role of competition in community dynamics. Oikos, **46**: 145–158.
Newman, M. 1995. Quantitative methods in aquatic ecotoxicology. Lewis Publishers, Boca Raton, Fla.
Newman, M. 1996. Ecotoxicology as a science. *In* Ecotoxicology: a hierarchical treatment. *Edited by* M.C. Newman and C.H. Jagoe. Lewis Publishers, Boca Raton, Fla. pp. 1–9.
Nisbett, R., and Ross, L. 1980. Human inference: strategies and shortcomings of social judgement. Prentice-Hall, Englewood Cliffs, N.J.
Peters, R.H. 1986. The role of prediction in limnology. Limnol. Oceanogr. **31**: 1143–1159.
Peters, R.H. 1988. Some general problems for ecology illustrated by food web theory. Ecology, **69**: 1673–1676.
Peters, R.H. 1990. Pathologies in limnology. Mem. Ist. Ital. Idrobiol. Dott. Marco de Marchi, **47**: 181–217.
Peters, R.H. 1991. A critique for ecology. Cambridge University Press, Cambridge, U.K.
Platt, J.R. 1964. Strong inference. Science (Washington, D.C.), **146**: 347–353.
Popper, K. 1968. The logic of scientific discovery. Hutchinson, London, U.K.
Popper, K. 1969. Conjectures and refutations. Routledge and Kegan Paul, London, U.K.
Quinn, J.F., and Dunham, A.E. 1983. On hypothesis testing in ecology and evolution. Am. Nat. **122**: 602–617.
Reckhow, K.H., and Chapra, S.C. 1983. Confirmation of water quality models. Ecol. Modell. **20**: 113–133.
Roughgarden, J. 1983. Competition and theory in community ecology. Am. Nat. **122**: 583–601.
Rousseau, D.L. 1992. Case studies in pathological science. Am. Sci. **80**: 54–63.
Salt, G.W. 1983. Roles: their limits and responsibilities in ecological and evolutionary research. Am. Nat. **122**: 697–705.
Shrader-Frechette, K.S., and McCoy, E. 1990. Theory reduction and explanation in ecology. Oikos, **58**: 109–114.
Shrader-Frechette, K.S., and McCoy, E.D. 1993. Method in ecology: strategies for conservation. Cambridge University Press, Cambridge, U.K.
Simberloff, D. 1983. Competition theory, hypothesis-testing, and other community ecological buzzwords. Am. Nat. **122**: 626–635.
Slobodkin, L.B. 1988. Intellectual problems of applied ecology. Bioscience, **38**: 337–342.
Suter, G.W., II. 1993. Ecological risk assessment. Lewis Publishers, Boca Raton, Fla.
Suter, G.W., II. 1995. Endpoints of interest at different levels