USE OF HYPOTHESES IN THE SCIENCE AND INFORMATION DIVISION: A RESPONSE

Nicholas Lander

(In the last issue of SIDNews, Ian Abbott discussed use of hypotheses)

I think that Lakatos, Kuhn, Feyerabend, Chalmers and others have written more than enough to falsify the statement in Ian's article published in *SID News 1(4)* claiming "a logically impeccable foundation" for the so-called hypothetico-deductive method! In my view, the acceptance of the hypothetico-deductive method conditioned by Popperian falsification for workaday purposes is rather akin to our acceptance of Newtonian mechanics rather than the complexity of quantum mechanics to deal with the exigencies of everyday life . At best a convenient aproximation, it doesn't bear close inspection: it is clearly a matter of cerebral economy; the logic against it is compelling.

Not all null hypotheses are subjected to the kind of statistical analysis suggested in Ian's article. In descriptive taxonomy, for instance, it is unusual (but by no means unheard of) to present such an analysis when splitting a taxon previously considered to represent a single species, a situation directly analagous to Ian's X and Y example. Perhaps it should be more widely applied. Rather, the argument hinges on a revision (or re-gathering) of the underlying observational data which is normally sufficiently dramatic to speak for itself without the need for statistical analysis as such. Witness Ian's own example of "The swans of Europe and Australia do not differ in colour": one hardly needs a t-test to falsify this null hypothesis.

Further, the claims of the falsificationsists are seriously undermined by the fact that some if not all observation statements are patently fallible. Which is to say that if a universal statement or complex of universal statements constituting a theory or part of a theory clashes with some observation, it may be the observation that is at fault. Nothing in the logic of the situation requires that it should always be the theory that is rejected. A fallible observation statement might be rejected and the fallible theory with which it clashes retained. Witness the conflict between the everyday observation that the moon is much larger when it is nearer the horizon than when it is high in the sky (still regarded as an unexplained illusion) with modern descriptions of the moon's trajectory. Chalmers (1988) devotes a whole chapter to the limitations of falsification which makes for very sobering reading..

I wonder if Ian's concern with the relationship between hypotheses and monitoring might not be broached via a brief consideration of the complimentary role of confirmation as apposed to falsification in science? This relates back to the twin notions of bold conjecture and novel prediction which Ian alludes to early in his article . Surely, a conjecture is more or less bold only when viewed against the background knowledge of the time. Further, the predictions of such a bold conjecture will necessarily be judged novel in that they involve some phenomenon that does not figure in, or is perhaps explicitly ruled out by the background knowledge of the time. Thus whenever a bold conjecture is confirmed cautious conjecture is necesarily falsified For this reason, *confirmations* of new theories are important in the growth of science in that they constitute evidence that a new theory is an improvement on the theory it replaces. All this is to say that insofar as monitoring serves the purpose of confirming the predictions of bold conjecture it, too, has its role to play.

The following fictitious example may serve to usefully illustrate some of the points I have made:

Hypothesis

A CALM plant taxonomist, through years of observation of species in the family in which she specialises, boldy conjectures that the well-accepted Genus A is polyphyletic, which is to say that as presently conceived it has more than one common ancestor and thus might more properly be classified into several genera.

Falsification of null hypothesis

To test the null hypothesis (ie that the putative genera are not different) our taxonomist resorts to the methodology of cladistic analysis, employing all available and relevant data for these species and for those in several putatively related genera to reconstruct the most likely lineage leading to the evolution of the present-day species. It is discovered that there is but one most parsiminous evolutionary tree consistent with the observed facts, ie one tree assumes fewer evolutionary steps to have occurred than any other possible tree. According to this model, of the species currently included in Genus A, some share a common ancestor with Genus X, others with Genus Y. What is more, the level of signficance placed on these results is acceptable, ie the cladogram is reasonably robust. Since these results patently contradict the null hypothesis under test, the taxonomist now formally segregates the 101 species of Genus A into two genera, namely Genus A and Genus B.

Confirmation of new theory

Although this new classification initially proves somewhat controversial, over the next ten years the author of what is now considered a landmark paper continues to monitor all new collections of the species in the genera concerned. During this time, a dozen or so new species have been described, all of which fall neatly into her classification, thus confirming the underlying hypothesis and gaining broad general acceptance for it.

Rejection of fallible observation statement and retention of fallible theory

A couple of years ago there was an interesting new development. A colleague working in London drew our taxonomist's attention to a small fragment of a specimen in the British Museum which would seem to represent a species intermediate between Genus A and Genus B. This fragment was collected in the nineteenth century on the area of land apparently now occupied by the R & I Tower in central Perth. No other specimens are extant, and it has not been re-collected from the metropolitan area or beyond. Thus it would seem reasonable to presume that the species concerned is now extinct. Understandably, our taxonomist's managers have been unwilling to allow her to revisit her studies of genera A, B, X and Y -- at least until she has completed her current revision of the distantly related Genus Z which comprises some 300 or so species, many of which have been found to contain powerful anti-viral agents potentially of immense economic importance.

Articles such as Ian's are so very difficult to write. On the one hand, there is a need for a simplified account of methodology that can be acted upon. On the other, one must avoid venturing too far into a specialised area without the necessary specialist equipment, thus running the risk of giving hostages to fortune either by loosing the attention of one's readers or (worse still) by exposing one's own inevitably limited knowledge of the terrain, which has certainly changed a bit since Popper's time.

In conclusion, might it not be better to draw the attention of CALM's scientists to that excellent little book by A.F. Chalmers, namely *What is this thing called Science? An assessment of the nature and status of science and its methods* published by University of Queensland Press (1988) and currently available at a cost of $17.95? Chalmers is readable, comprehensive and balanced. He covers everthing from induction, through theory-dependent observation, falsification (Popper), theories as structures (Lakatos' Research Programmes, Khun's Paradigms), objectivism, anarchism (Feyerabend), realism, instrumentalism and truth, unrepresentative realism, and many other "isms".

Perhaps the more interesting issues raised by the taxonomic parable above are managerial rather than scientific in nature.

Comment on "Use of hypotheses in Science and Information Division" by Ian Abbott

Matthew Williams

I enjoyed this article , and my only reservation about it is that it was too short. Because of this, the article oversimplified some aspects of the scientific method relating to hypothesis testing. In my comments below I try to clarify some of these aspects:

1. The article argues that the hypethetico-deductive method is the only method of science. This is wrong. Sir Ronald Fisher also stated (Fisher (1935): "The Design of Experiments") that the purpose of experimentation is to test hypotheses, and he was wrong too. Consider, for example, an experiment designed to compare five different tree species in their production of some commodity, say the amount of millable timber each produces after ten years. In a typical experiment, the five species would be planted together in a number of replicates, perhaps at a number of sites. The yield of timber would then be compared, and analysed, probably by ANOVA. The implicit null hypothesis underlying such an ANOVA is "(X): All five species produced the same amount of millable timber (i.e. the mean yields are all equal)" (this could be tested by the F statistic). However, nobody would ever believe such a null, and the purpose of this null is merely to provide a framework for the experiment. We do not even need to look at the F statistic, other than to assess the probability that any observed differences were due to chance. The real purpose of this experiment was "(Y): to objectively determine the best producing species, and to estimate its yield". It seems to me that to state hypothesis (X) as part of the aim of this experiment would be unneccesary (and indeed misleading), wheras (Y) should be stated as part of the aim.

2. Another point that I think needs clarification is the distinction between hypothesis testing and decision making. These two tasks, the former scientific and the latter political, are quite separate. In the discussion of the two alternate hypotheses (A: "Prescribed burning of jarrah forest in spring does not spread *Phytophthora* fungus"), and (B: "Prescribed burning of jarrah forest in spring causes the spread of *Phytophthora* fungus"), it is stated that "If the evidence collected does not refute (A), then we do not need to consider (B)". I agree that it is nonsensical to attempt to test (B), since (A) is a null hypothesis and hence more amenable to disproof. However, I think there are two points which should be noted:

(i) To assume that (A) is true, until disproved, is to confuse the artifice of the hypethetico-deductive technique with common sense. Before we test (A), we do not know if *Phytophthora* fungus is spread by spring burning. We merely propose (A) (as a null hypothesis) because it is a (relatively) easy statement to test using the hypethetico-deductive technique. [As an aside, how we choose to manage the forest until (A) is proved or disproved is a purely political decision - it may be prudent, in order to protect the forest resource, to assume that (B) is true and act accordingly.] If the evidence collected does not refute (A), then we should still consider (B). We must always bear in mind that if our experiment is inadequate (for example, too small), then we may never be able to refute (A). In the final decision of whether (A) or (B) is true, we need to consider the weight of evidence relating to (A) and (B). If a very small experiment fails to reject (A), our chance of making a type II error (i.e. accepting (A) as true when in fact it is false) is high. Conversely, this chance is low if a very large/comprehensive/powerful experiment was performed. [This fact has lead to certain environmental protection agencies in the US requiring that in environmental impact studies, the power of the experiment designed to detect an impact, is specified. Otherwise, the outcome could be "rigged" to provide a result of no impact, simply by doing a very small study.]

In conclusion, in making decisions about the jarrah forest, managers need to consider both (A) and (B), and the chance that each is true. Thus it is perfectly defensible for managers to assume that (B) is true, if they adjudge that the evidence