Statistics as Principled Argument

Robert P. Abelson

Mahwah, NJ. Eribaum Associates, 1995, 0-8058-0528-1, $32.50US (paper), 232 pp.

Reviewed by Peter McBurney

The economist Milton Friedman once remarked: "the role of statistics is not to discover truth. The role of statistics is to resolve disagreements between people." (quoted on p. 4 of McCloskey 1994). Robert Abelson's book, Statistics as Principled Argument, takes this view of statistics and uses it as an organizing principle for explaining the use of statistical methods. The book is aimed at beginning graduate students or advanced undergraduates in the social sciences, and is based on the author's 35 years teaching introductory statistics in the Psychology Department at Yale.

The book is a novel introduction to statistical methods and is replete with fascinating examples and case studies from the social sciences research literature. I believe these examples would best be appreciated by students who had already taken a first statistics course, and thus already have some familiarity with the methods described. Hence, the book could be used as a text for a second course in statistics or for a graduate seminar in research methods. Graduate students in other experimental sciences, such as those in biology and medicine, as well as those in the social sciences, would also gain from working their way through this book. In addition, applied statisticians whose work involves assisting or working with social or biomedical scientists may benefit from it.

The author argues that four criteria are crucial to assessing the persuasiveness of statistical arguments in social science domains. First is the magnitude of any experimental effect(s) observed. Differences which are highly statistically significant are more persuasive than ones which are less significant or not at all significant. Second, is what Abelson calls articulation, which refers to the level of detail at which scientific results are presented. Results expressed in quantitative terms typically have more impact than those expressed only qualitatively, and, of these, more specific claims have greater impact than less specific. Third, is the generality of effects observed, by which Abelson appears to mean the extent to which observed effects are replicable and generalizable beyond the particular experimental context in which they were observed. Fourth, is the interestingness of claims, which is a combination of their significance for current theories and their degree of surprisingness. The

more general and the more interesting is a claim, the greater, typically, is its impact on an academic community. Abelson then argues, based on his experience, that to be accepted into the lore of an academic community, a research paper must satisfy at least three of these four criteria.

Why should any of this interest philosophers of argumentation? I think there are two reasons. The first arises from the fact that Abelson's criteria for persuasive statistical argument are based on his long experience as an academic researcher and teacher in psychology; they are not grounded in any theory of belief revision in the social sciences, nor based on any explicit philosophy of social or experimental science. One is tempted to ask: Why these criteria and not others? Why must a research paper score highly on at least three criteria? What happens to research papers which score highly on only two? What happens to two papers with equally high scores but with conflicting results? etc. I believe that adequate answers to these questions would require an argumentation-theoretic account of community knowledge creation and revision in the experimental sciences. To my knowledge, no such account exists. Indeed, the philosophy of science in general has yet no adequate, argumentation-theoretic account of knowledge creation and revision, although several authors have taken first steps in this direction (e.g., Pera 1994, Goldman 1999 and, in the philosophy of mathematics, Kitcher 1984). I believe such an account would be a worthy goal in and of itself, as it would cast light on how scientific knowledge accumulates and changes and, perhaps, how it should do so.

In addition, an argumentation-theoretic account could also provide practical advice to scientists in particular situations. In the final chapter of his book, Abelson discusses an actual case involving a research paper claiming to show that group meditation led to a reduction in crime rates, auto accidents and fires (Orme-Johnson et aL 1988). There is no currently-accepted causal mechanism in western science which links meditation to these phenomena. Consequently, Abelson argues that an editor of a journal to which the paper is submitted would face a difficult decision in deciding to accept or reject it for publication: "if the study appears to be methodologically sound, it would be contrary to the ideal of scientific openness to give it no hearing whatever, no matter how seemingly cockamamie its claims. The editor has a problem." (p. 179). I believe Abelson to be profoundly wrong on this point. The purpose of western science, surely, is the development of explanations of observed natural phenomena, not the identification of natural phenomena to exemplify prior established explanations. If an experiment identifies some important phenomena for which no scientific causal explanation yet exists, this itself is interesting and deserving of publication. The only criteria for assessment for publication should be the statistical rigor with which the study was undertaken and the importance of the phenomena. There would be no science as we know it without such publication. For instance, the entire field of environmental and

medical risk analysis arguably works in this way: a possible correlation is observed between some behavior, environment or substance (such as thalidomide) and its effects (such as deformities in offspring), and then studied experimentally. Only after the effect is observed are possible causal mechanisms typically proposed for the phenomenon, and often by people different from those who first observed the correlation. The burden of finding acceptable causal mechanisms does not lie with those who first observe the phenomena, but with the entire scientific community or communities whose domain is involved.

There is a second reason why this book should interest philosophers of argumentation. If statistics is a form of argument in experimental sciences, why are some, particular, statistical methods acceptable as arguments rather than others. In other words, what are the arguments for using particular statistical arguments? This is a question the book does not consider. An adequate answer would involve an understanding of the sociology, philosophy and history of the social & biomedical sciences, in addition to an understanding of the reasons which statisticians, in their own literature, advance for preferring one technique over alternatives. These reasons usually involve sophisticated pure mathematics—e.g., aspects of measure theory, differential geometry, and algebraic topology—most of which is well beyond the mathematical expertise of social and biomedical scientists. However, not all the reasons involve deep mathematics. For example, the arguments for setting significance levels in tests of statistical hypotheses at the standard levels (5% or 1% for the probability of wrongly rejecting a true null hypothesis) are based on informal considerations which arose in the context of agricultural experiments in the 1920s and 1930s; their justification is certainly not rationally uncontestable, and nor are these levels necessarily appropriate in other domains, as I argue in a forthcoming paper (McBurney & Parsons 2002). An argumentation-theoretic account of the accumulation of statistical knowledge would also be a nice theory to have. It too may have practical import as discussions over the use of the Precautionary Principle in risk assessment reveal, since these are essentially debates over the appropriate values for levels of significance in hypothesis tests.

In conclusion, this book is well-written, interesting and novel, and its intended audience of beginning researchers in the social sciences should find it of great value Moreover, it points to a set of important research questions in the philosophy and practice of the experimental sciences whose resolution would require major contributions from argumentation theorists.

References:

Goldman, A. (1999): *Knowledge in a Social World*. Oxford, UK: Clarendon Press.

Kitcher, P. (1984): *The Nature of Mathematical Knowledge*. New York, NY, USA: Oxford University Press.

McBurney, P. and S. Parsons (2002): "On hypothesis testing procedures in risk assessment: a research agenda." *Law, Probability and Risk* (in press).

McCloskey, D.N. (1994): *Knowledge and Persuasion in Economics*. Cambridge, UK: Cambridge University Press.

Orme-Johnson, D.W. and C. N. Alexander, J. L. Davies, H. M. Chandler and W. E. Larimore (1988): "International peace project in the Middle East: The effects of Maharishi technology of the unified field." *Journal of Conflict Resolution*, 32: 776-812.

Pera, M. (1994): *The Discourses of Science*. Chicago, IL: University of Chicago Press.

Peter McBurney
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF
United Kingdom
p.j.mcburney@csc.liv.ac.uk