

Book Reviews

Books for review are to be sent to the Book Review Editor Jaki Stanley, USDA/NASS, Research Division, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A.

Abelson, R.P., <i>Statistics as Principled Argument</i>	317
<i>Oyvind Hoveid</i>	
Bartholomew, D.J., <i>The Statistical Approach to Social Measurement</i>	319
<i>Jean-Christophe Bureau</i>	
Schwarz, N., <i>Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation</i>	321
<i>Janet Harkness</i>	

Abelson, Robert P., *Statistics as Principled Argument*. Hillsdale, NJ:LEA, 1995. ISBN 0–8058–0528–1, 198pp. +refs. and index, 22.00USD.

The author of *Statistics as Principled Argument* is both a psychologist and a statistician. Accordingly, this book draws heavily upon his insights in psychology. This enables him to provide insights in statistics that are rare, or at least not clearly expressed, among pure statisticians. One might say that the book deals with the rhetoric of statistics, or rather meta-statistics.

Numerous cases from psychology illustrate Abelson’s points, which suits the text well. The reader is prevented from viewing statistics as the art of throwing dice, playing roulette, or testing hypotheses in formal models. Abelson’s major concern is empirical research: How and when can experiments and observations combined with statistics reveal insights in real life? In this way he considers statistics as a tool – not a discipline in its own right. His comparison of statistics and English is illustrative. Although lies and half-truths are expressed in English, one does not discard the language because of that. In fact, knowledge of both English and the subject is required to gain important insights.

Abelson begins by acknowledging the conceptual difficulties of statistics. He says three pass-throughs of statistics are necessary. One for exposure, one for practice, and one for the dawn of genuine insight. The book is definitely for the last pass-through. At first sight the text seems elementary as formulae with all those Greek letters and indices do not appear, and he makes absolutely no effort to include technical details. Nevertheless, he gives an insight into the counterintuitive nature of chance, and he makes it perfectly clear what hypothesis testing, p -values etc., are about and why these tools are useful. The reader who is not familiar with statistics at the technical level, will perhaps not quite understand. I believe the informal text is more useful, the deeper the reader is into formal statistics.

The book is composed as an explanation of the so-called MAGIC criteria – magnitude,

articulation, generality, interestingness, and credibility – which an argument based on statistical inference should satisfy. With respect to magnitude, Abelson is not perfectly clear. He mentions both statistical test values and effect values as important, but does not relate them too much. He might have mentioned that a high effect value with dubious statistics is a call for more research to get the effect confirmed or rejected on more solid grounds. Extreme test values with small effects are on the other hand usually not very interesting except for their indication of errors.

Regarding styles of rhetoric, he explains various positions along an axis from brash through liberal and conservative to stuffy. He shows that suitable selection of wording accompanying statistical tests, the consistency or inconsistency between an hypothesis and statistical material can be presented with different degrees of strength. At this point, I find Abelson too informal. Formal statistics have more to contribute at this point. Tests should be chosen because of their applicability, not their outcome. If outcomes are different, then the tests are based on different assumptions which in turn can be subject to testing or reasoning on which test to choose. This does not totally eliminate the role of styles, though, but the scope is diminished.

Abelson has an important chapter on fishiness. In the age of computers the ability to produce systematic errors has perhaps increased more than the ability to discover them. Therefore, special attention aimed at detecting errors is required. Usually one does not find an error without having searched for it. Thus, his hints on indications of errors are extremely important.

When it comes to articulation and generality I find that his text is somewhat hampered by his background from the social sciences. This has lead him to concentrate on the prevalent two-by-two tables. He shows convincingly that a redefinition of columns and/or rows may make a great difference to the statistical significance. Abelson states that it is important to find suitable columns and rows leading to claims (the distinct research results or ‘ticks’) which are supported by the data. Findings that certain claims hold in certain situations but not in others (the statements that qualify the ticks or ‘buts’) may also be interesting in their own right. Tables without a convincing story (undifferentiated research results or ‘blobs’) should be avoided. To me, a table design is nothing but a simple non-parametric statistical model of the observations. Experiments can be designed to fit with m -by- n tables. Otherwise, there are numerous alternatives, both non-parametric and parametric to the m -by- n design. He does give a warning against heedless search for significant tables, but he might have been more clear on this point.

The text recovers greatly in the chapters on interestingness and credibility. This is more about scientific arguments than statistics. Here he relates statistical claims to theory and conventional wisdom, and he points to two interesting games to play. Either one should gather contradictions to what is currently accepted, or find new tenable positions where ignorance and contradictions so far have ruled. In both cases, there may arise arguments to be won or lost. Those who have read Abelson’s insights in the psychology of scientific arguments will certainly be favored in these arguments.

The author excuses his age and use of old examples as illustrations. However, his age and experience have allowed him to gather a considerable number of amusing and informative examples. The book is just as much about psychology as statistics and this makes it highly recommendable. ‘Abelson’s Laws,’ informal rules regarding

statistical rhetoric, are woven throughout the book and are a major contribution worth the price of the book alone.

Oyvind Hoveid
Norwegian Agricultural Economics Research Institute
PO Box 8024, N-0030 Oslo, Norway
Phone: ++47 2217 3540
Fax: ++47 2217 3538

Bartholomew, David J., *The statistical approach to social measurement*. Academic Press Inc., San Diego, CA, U.S.A. 1996. ISBN 0-12-079860-3, XII+239 pp., 54.95USD.

The aim of this book is to provide “a unified statistical approach to the measurement of social phenomena.” Here, the term social is used in the broadest sense. It includes psychology, demography, and sociology as well as business studies, and the different chapters are concerned with measures of mobility, inequality, and individual characteristics and performances. The term statistical approach, on the other hand, is used in a relatively narrow sense, since the book focuses on stochastic approaches, namely on the statistical properties of the implicit models that underlie empirical measures. Buyers should not expect to find a standard or detailed presentation of, say, classical index number theory, national accounting techniques, or educational testing indexes. Many of these indexes are described in the book. However, they are presented in a rather academic way as particular cases of a general statistical framework. Instead, readers will find the author revisiting the standard techniques by using stochastic approaches, with an emphasis on latent variables analysis.

The common point raised in different issues throughout the book is that measurement is derived from a representation of problems using random variables, and that measures rely on parameter estimates in a statistical model. This makes it possible to clarify the criteria by which a measure should be judged, an issue investigated in Chapter 2. By linking measurement to statistical modelling, the concepts of validity and reliability of a measure can be related to precise statistical concepts.

Chapters 3 to 6 deal with some examples of population level measures. Chapter 3 is concerned with labour wastage measurement, i.e., the loss of individuals from a firm. Wastage is treated as a stochastic process developing in time. Focusing on the inner dynamics that drive the process, the author develops a statistical approach where the key element is the identification of a hazard function. Chapter 4 deals with inequality measurement. Rather than motivating the choice of an inequality index on the basis of a social welfare function, or on an axiomatic exploration of the compatibility of the index with basic properties required to satisfy philosophical principles, the author considers that income distributions are the outcome of a stochastic process, i.e., people moving up and down the income ladder, and focuses on the relative dispersion of a frequency distribution. He investigates the statistical foundations of indicators, including the most common ones, such as the Gini coefficient.

The whole philosophy adopted in the book is that measures should be derived by reference to a model of the process being studied. The goal is to identify some parameters as encapsulating the concept in question. An estimate of the parameter is considered as the

relevant measure. This is particularly explicit in Chapter 5, dealing with mobility measurement, and in Chapter 6, dealing with price level measurement. In Chapter 5, Markov chains are emphasised as models for mobility. In Chapter 6, price levels are identified as a function of time that summarises the individual price curves, and an error term captures the fact that all prices do not follow the same variation. Price indexes can therefore be estimated by regression. Various specifications, including Theil's approach and Fisk's approach are described in detail.

In Chapters 7 to 10, attention shifts from aggregate behaviour to the measurement of individual characteristics such as abilities and attitudes of individuals or firms. The author's main assumption is that the numerous determinants of individual behaviour should be treated like any other variable, although they cannot be directly measured. Their measurement is therefore not essentially different from the population measures based on parameters, even though individual variables are latent variables, and estimation involves specific techniques. Indicators provide the link between the world of things that can be measured and the underlying world of latent variables. Measurement therefore requires finding a relevant set of indicators and a methodology for constructing a scalar measure of the latent variable from those indicators. Note that it is sometimes possible to use data analysis techniques which do not require invoking a specific model.

Chapter 8 is based on the assumption that the items used to elicit a person's scale position can be located on the same latent scale as the individuals themselves. In practice, for example, the measurement of attitude, ability, or preferences can be approached by an ordered set of questions, that express the same basic attitude in stronger or weaker form. Chapter 8 provides a detailed survey of methods of scaling, including the Guttman scaling and Rasch's model. Chapter 9 generalises these methods and provides a systematic treatment of the general problem of measurement at the individual level, provided that there exists a continuous scale on which individuals can be located. The general framework is to specify a form for the joint probability of the individual responses and the scale point, considered as random variables, and to extract the information from an individual's response based on the expected value of the conditional distribution of the observed response to the latent variable.

Chapter 10 is mainly a set of examples to show how the general ideas expressed in the previous chapters work out in practice. Many illustrations are provided in the field of ability testing, measurement of attitudes, clinical depression and social needs. Empirical applications are detailed, and the author indicates how to implement the methods described on a computer, with the appropriate software. Chapter 11 draws attention to some of the limitations inherent in social measurement, and identifies directions in which new research is required.

The main strength of this book is to provide a general framework to measurement problems, using statistical modelling. In many chapters, the demonstration that various widely-known techniques are nothing but different aspects of this general approach is compelling. This is particularly clear in the chapter on price levels, where the author provides a statistical foundation for the Divisia index, and as a result, for the commonly used Laspeyres, Paasche and Fisher indexes. Although his presentation of both the economic and axiomatic approach of index numbers is very brief, the author manages to make it clear that linkages exist between the three approaches, and that beyond specialised fields of the literature, there are common foundations.

Perhaps, the reader may not share the idea, implicit throughout the book that the author's statistical approach presented is more general than other approaches, and therefore superior to them. For example, one may consider that the economists' approach to equality measurement, relying on a social welfare function, is too easily dismissed by the author in Section 4.6. In the same way, the economic approach to index numbers in price level measurement has some internal consistency and rigour even without reference to the statistical properties of an underlying model (see Diewert, for example).

However, there is no question that the book is an important step toward the "Great Unified Theory" of social measurement and that Bartholomew manages to bring into perspective different fields of the literature in a very clear and compelling way. The derivation of many empirical properties and practical recommendations from the statistical properties of the underlying model is impressive. Note, however, that this book is intended mainly for academic researchers and graduate students. Practitioners will not find accessible ready-to-use tools, but those who have enough knowledge in statistical inference and probability will find in it a theoretical framework within which their needs for new tools can be met.

Jean-Christophe Bureau
Institut National de la Recherche Agronomique
Département d'Economie
78850 Grignon, France.
email: bureau@grignon.inra.fr
Phone: +33 (0) 130815336
Fax: +33(0) 130815368

Reference

Diewert, E., (1995). *Essays in index number theory*. North Holland.

Norbert Schwarz., *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Lawrence Erlbaum Associates, 1996. ISBN 0-8058-2314-X, 91pp. + ref. and index, 22.50USD.

This slim volume of 100 odd pages can be seen as the thin edge of a rather welcome wedge or, depending on your frame of reference, as a slender foot-in-the-door. The book is based on a series of lectures delivered as the 1995 John M. MacEachran Memorial lectures at the University of Alberta and on material in Schwarz (1994), a publication incidentally mentioned in the acknowledgments but absent from the references.

Books and lectures are composed and presented with audiences in mind. The audience at a memorial lecture series in honour of an eminent psychologist is likely to have been conversant with the goals, techniques, and terminology of experimental psychology but perhaps less familiar with the field of cognitive psychology in survey research (here *CPSR*), the field to which Schwarz has made significant contributions since the early eighties.

The contents of the book are, in fact, relevant for a variety of audiences, but particularly for researchers interested in understanding findings on respondents' responses to questionnaire and test items. The ideas presented within a framework of Gricean conversational

norms and principles are of importance for instrument design, data analysis, for understanding respondent behaviour and, we might add, understanding respondent needs.

Readers are provided with a condensed overview of cognitive research on response effects of diverse kinds undertaken since the early eighties. Schwarz outlines the conclusions hitherto drawn about respondent judgement potential on the basis of these findings, then presents his own perspective, based on what he calls a conversational approach to questionnaires. Literature references are numerous (a twelfth of the book), many themselves including literature reviews. Thus the volume can usefully serve as an advanced overview of the CPSR work in which Schwarz and colleagues have been engaged and, to a certain extent, the effect this has had on survey research. The short, clear presentations are valuable to anyone needing a quick review (or re-view) of this research.

However, this reviewing aspect of the book is only a means towards an end. The main purpose is to take stock of CPSR findings and to (re)-interpret them from a 'conversational' perspective. A central CPSR contribution to survey methodology research has been to demonstrate a logic behind survey response effects. In this connection, researchers such as Schwarz have referred to the Gricean maxims of conversation and the Principle of Co-operation, which contribute to shaping how, what and why we communicate what we do in given contexts. Here, however, Schwarz shifts and sharpens the focus. Conversational logic involves more than truth value logic. In conducting conversations, we rely extensively on deductive inferences to arrive at our interpretations of what communications mean in a given context. In inferring, we utilise everything we perceive to have communicative import – context, co-text, knowledge of the world and social norms, together with a refined screening of what is said, by whom, and how. In deciding on the meaning of what it is they are being asked, respondents proceed in a similar manner, processing everything they perceive as relevant to the discourse in order to best understand what is required by way of a response.

Surveys, Schwarz proposes, are conversations in which the respondent side is consistently co-operative and observes conversational maxims. Respondents regularly utilise ('read') elements of questionnaires neglected (or 'unread') by researchers in order to co-operate as best they can in responding. The other participant in the conversation, the researcher side, is seen as non-co-operative, as frequently disregarding these maxims, either deliberately to pursue research interests or through failing to recognise the discourse character of the interchange. The interaction of these two components – involvement in a 'conversation' and the asymmetric observance and non-observance of conversational maxims – are, Schwarz suggests, the framework within which findings on respondents' judgmental biases and the so-seen 'error proneness' of responses are to be viewed. Simply put, if this is accepted and researchers recognise what respondents may perceive, responses and judgements start to make more sense and better logic.

The book is divided into seven chapters. The introductory chapter outlines the contents of the other chapters and presents the main thesis of the volume – that principles which govern everyday conversation are centrally relevant in understanding findings in judgmental research.

The second chapter outlines the maxims of conversational practice advanced by the language philosopher Grice. Together with a number of other authors in language philosophy, linguistics and different fields and schools of text analysis, Grice's ideas have

played a key role in changing our understanding of meaning(s) and communication, literally revolutionising areas of semantics, discourse theory and text analysis (not dealt with in the volume).

In the third and fourth chapters, Schwarz reviews findings from cognitive research on perceptions and judgements which show predictable effects, depending on, for example, *what* people are presented with, the *order* in which things are presented and the intentional or unintentional *additional information* provided alongside these elements.

Chapter five deals with the effect of certain so-called 'formal aspects' of surveys on respondents' perceptions and judgements, focusing in particular on response scales. Various systematic effects observed in connection with specific features of response scales are used as illustrations of respondents' co-operative communicative behaviour. The chapter also discusses respondent perceptions of what is *perceived as asked* in contrast to what is *intended to be asked*, a central distinction in modern discourse and text theories. It illustrates how perception can be predictably manipulated through question type (open versus closed), through question sequences (general followed by specific, specific followed by general), through pairs of questions versus larger sets, the provision of filter and quasi-filter questions or, for example, by offering a *don't know* category.

Chapter six focuses on the communicative patterning or conversational logic behind such findings. Parallels are drawn between interpretative behaviour as exhibited by adult respondents and findings on children's perceptions of researcher communicative intent. In both instances, findings taken as evidence of people's inability to perceive correctly are re-interpreted as evidence of how well and holistically people actually perceive in terms of the conventions of conversational logic.

The last chapter concludes firstly '*that the key methodological implication of a conversational analysis of research procedures is that researchers need to be sensitive to the information that their procedures provide to their participants*' and secondly (as the concluding sentence) that '*(t)aking the tacit assumptions which govern the conduct of conversation into account provides a promising start for this endeavour.*'

As someone working on questionnaires as text and discourse, I welcome both conclusions and would like to add a few signposts from a text and discourse standpoint. Rather than focusing on *conversational* discourse, I think we need to develop a general discourse and text perception of questionnaires and of respondent-researcher interaction via questionnaires. Ultimately, we can bring together research on text types and text comprehension, on discourse types and discourse interaction, on CPSR research as presented here, and work on the visual and graphic aspects of questionnaires (e.g., Jenkins and Dillman 1997). I think, too, that issues of researcher co-operation or lack of co-operation and the flouting of maxims on both sides are an important area for future research.

Questionnaires are texts peculiarly destined for discourse (Harkness 1994) and belong to highly structured discourse and text types. In closed question formats, even respondent 'answers' are pre-formulated (Harkness 1995). In processing questionnaire texts, respondents may well utilise text strategies relevant for tax forms or license applications, as well as those for less technical question-and-answer discourse situations. Narrative structures, textual relationships and persona (Harkness 1996) shape respondents' perceptions of both themselves and researchers. While tacit assumptions governing conversational interaction are relevant in questionnaire discourse, conversational analysis (CA) only deals with

certain aspects of this discourse. (See, incidentally, Hootkoop-Seenestra 1995 and Schaeffer, Maynard, and Cradock 1995 for a CA approach to oral interview data.) A discourse and text standpoint would require researchers and questionnaire designers to see questionnaires as discourse entities, to recognise that textual components present (or absent) are given readings, and, one could hope, to commit themselves not just to respondent-friendly graphic presentations but to respondent-acknowledging discourse.

I began by describing the book as the (welcome) thin edge of a wedge. Views on whether we have before us a door, a frontier, or a Pandora's box will probably differ before communication-related research is properly incorporated into survey research.

References

- Harkness, J. (1994). Who is Who in Questionnaires. Paper presented at the World Congress of Sociology, Bielefeld, July.
- Harkness, J. (1995). Getting a Word in Edgeways: Questionnaires as Text and Discourse. Paper presented at the annual meeting of the American Association for Public Opinion Research, Ft. Lauderdale, FL, May.
- Harkness, J. (1996). The Representation of Self in Everyday Questionnaires. Paper presented at the ISA Conference on Survey Methodology, Colchester, U.K., June.
- Houtkoop-Steenstra, H. (1995). Conversational Problems in Telephone Interviews. Paper presented at the International Conference on Survey Measurement and Process Quality, Bristol, U.K., April.
- Jenkins, C. and Dillman, D.A. (1997). Towards a Theory of Self-Administered Questionnaire Design. In *Survey Measurement and Process Quality*, Lars Lyberg et al. (eds.). New York: John Wiley and Sons.
- Schaeffer, N.C., Maynard, D.W., and Cradock, R.M. (1995). Negotiating Uncertainty: Uncertainty Proposals and Their Disposal in Standardised Interviews. Paper presented at the International Conference on Survey Measurement and Process Quality, Bristol, U.K., April.
- Schwarz, N. (1994). Judgement in a Social Context: Biases, Shortcomings, and the Logic of Conversation. In *Advances in Experimental Social Psychology*, M. Zanna (ed.), Vol. 26, pp. 123–162. San Diego, CA: Academic Press.

Janet Harkness
ZUMA
PO Box 122155
68072 Mannheim, Germany
e-mail: harkness@zuma-mannheim.de
Phone int + 621-1246-284
Fax int + 621-1246-100