Research Methodology

in the Social, Behavioural & Life Sciences

Edited by Herman J Adèr
& Gideon J Mellenbergh

RESEARCH METHODOLOGY

RESEARCH METHODOLOGY

in the Life, Behavioural and Social Sciences

Edited by
Herman J. Adèr and Gideon J. Mellenbergh

SAGE Publications
London • Thousand Oaks • New Delhi

First published 1999

SAGE Publications Ltd
6 Bonhill Street
London EC2A 4PU

SAGE Publications Inc
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
32, M-Block Market
Greater Kailash - I
New Delhi 110 048

British Library Cataloguing in Publication data

A catalogue record for this book is
available from the British Library

ISBN 0-7619-5883-5
ISBN 0-7619-5884-3 (pbk)

Library of Congress catalog record available

Printed and bound in Great Britain
by Athenaeum Press, Gateshead

Contents

List of Figures

List of Contributors

Herman J. Adèr is a senior staff member of the department of Clinical Epidemiology and Biostatistics at Free University, Amsterdam. He studied Mathematics and Social Science Methodology at the University of Amsterdam. He has been a statistical advisor for many years, first at the psychology department, later at the faculty of Medicine of Free University. In these functions he contributed to numerous research articles and several PhD theses.

In his PhD thesis a formal system to represent methodological knowledge is proposed. Current research interests include graphical knowledge representation, model search methods and neural network modelling in which fields he published several papers and articles.

Gideon J. Mellenbergh (1938) studied psychology at the University of Amsterdam. He received his Masters Degree in 1965 and his PhD Degree in 1971 (English translation of the Dutch dissertation title: 'Studies in achievement testing'; chairman of the dissertation committee: Prof. Dr Adriaan D. de Groot). Since 1982 he is professor of Psychological Methods at the University of Amsterdam. Moreover, he is director of the Interuniversity Graduate School of Psychometrics and Sociometrics (IOPS), associate editor of Psychometrika, and member of the editorial board of Psychological Methods. His scientific interests are in fundamental and applied psychometrics, especially construction and analysis of psychological and educational measurement instruments, differential item functioning, psychometric decision-making, and theoretical analysis of psychometrical concepts and models.

Jaap van Heerden is Associate Professor in Methodology and the Foundations of Psychology at the University of Amsterdam. He contributed as author and co-author to the Journal of Mathematical Psychology, Inquiry, the Journal for the Theory of Social Behaviour, and New Ideas in Psychology. His scientific interest at the moment concerns consciousness studies.

H.P. (Paul) de Greef studied psychology at the University of Amsterdam. In 1984 he joined the department 'SWI' of the University of Amsterdam. This department provided a meeting ground for computer science, social science and behavioural sciences. In the group of Prof. Wielinga he worked in various European research projects on knowledge acquisition and representation. In 1989 he was appointed lecturer in human-computer interaction at a similar SWI department of the University of Leiden. There he obtained a number of large research grants. In 1991 the SWI department in Leiden was terminated and he returned to Amsterdam. His current interests are in representation, modelling and cognitive engineering. In 1999 he was awarded a PhD on a thesis entitled: 'Computer support by knowledge enhancement: constraints and methodology'.

Paul Darius studied mathematics at the Katholieke Universiteit Leuven, where he obtained his doctorate in 1982. He teaches statistics at the K.U. Leuven faculty of agricultural and applied biological sciences, and is currently director of the International Study Program in Statistics there. His research interests include mixed model methodology, meta data, statistical expert systems, visual software for experimental design and software for teaching statistics.

Kenneth M. Portier is an Associate Professor of Statistics and Agricultural Experiment Station Statistician at the University of Florida in Gainesville Florida. A 1979 graduate of the Biostatistics program at the University of North Carolina at Chapel Hill, Dr Portier has worked as a statistical consultant on agricultural, ecological and environmental research projects in the Southern United States, Central and South America. He has taught courses in statistical methods for research, natural resource sampling and advanced statistics for environmental biologists. His publications cover a wide range of statistical applications with more recent work on computational methods in statistics.

Stuart J. Pocock has been a Professor of Medical Statistics at the Medical Statistics Unit at the London School of Hygiene and Tropical Medicine since 1989. He received his MSc at Imperial College of Science, Technology and Medicine in 1968 and his PhD at the London School of Hygiene and Tropical Medicine in 1972. He is the author of the well-known textbook 'Clinical Trials, a practical approach' published by Wiley. His present interests are in standards for reporting on randomized controlled trials, statistical consultation, stopping rules for clinical trials, and Meta-analysis.

Jelke G. Bethlehem studied mathematical statistics at the University of Amsterdam. After obtaining his pre-doctoral degree he was employed as Research Worker at the Statistical Department of the Mathematical Centre in Amsterdam. His work concentrated on multivariate statistical analysis and development of statistical software.

In 1978 he joined the Department for Statistical Methods of Statistics Netherlands, first as Research Worker and later as Senior Statistician. His main topics were the treatment of non-response in sample surveys, in which he obtained his PhD, and disclosure control of published survey data.

From 1987 to 1996 he was head of the Statistical Informatics Department. The work of this department concentrated on the development of standard software for processing survey data. Important projects were the

Blaise System for computer-assisted survey data collection, the StatLine system for the electronic dissemination of statistical publications, and the Bascula package for weighting sample survey data.

Now he is senior methodogical adviser of the the Statistical Methods Department. He is involved in several research projects in the field of survey methodology. He is also part-time professor in Statistical Information Processing at the University of Amsterdam.

Peter C.M. Molenaar's areas of statistical experience include dynamic factor analysis, applied non-linear dynamics, adaptive filtering techniques, spectrum analysis, psychophysiological signal analysis, artificial neural network modelling, covariance structure modelling and behaviour genetical modelling. He is a Professor of Psychology at the University of Amsterdam. He published widely in the above mentioned areas, emphasizing applications to cognitive development (stochastic catastrophe analysis of stage transitions), brain-behavior relationships (real-time artificial neural network simulation of cognitive information processing), brain-maturation and cognition (equivalent dipole modelling of longitudinal registrations of electrocortical fields), genetical influences on EEG during the life span, and optimal control of psychotherapeutic processes. He teaches courses on covariance structure modelling, dynamical factor analysis and state-space modelling, psychophysiological signal analysis, and artificial neural network modelling. He has been and is the Principal Investigator of 21 grants focusing on methodological issues from the Dutch Scientific Organization, the Royal Dutch Academy of Sciences, and the University of Amsterdam.

Joe Whittaker is a Senior Lecturer in the Department of Mathematics and Statistics, Lancaster University. He is a Fellow of the Royal Statistical Society, past member of its Council, Research and chairperson of its Multivariate Study Group. He has been an elected member to the Board of Directors of the European Region of the International Association for Statistical Computing. He has held visiting Professorships in the Department of Statistics at Colorado State University, University of Chicago, and Paul Sabatier University, Toulouse. He is the author of over 60 articles in the open research literature. His research interests lie in graphical modelling, transport models and credit scoring.

Whittaker's contributions include his international best selling graduate text on graphical models. He has given numerous short courses on graphical models some sponsored by the Economic and Social Research Council, U.K. and have been given in Europe and America. He was co-organizer of the conference on Probabilistic Graphical Models, held at the Isaac Newton Institute, Cambridge, part of the Neural Networks 1997 Program in which he was a Visiting Fellow. He was the Program Chair for Uncertainty99, the Seventh International Workshop on Artificial Intelligence and Statistics, held in Florida 1999.

Pete Sewart obtained his first degree in Mathematics and Operational Research from Lancaster University. He continued his studies at Lancaster for his PhD which was supervised by Joe Whittaker, concentrating his research on graphical and longitudinal data models relevant to the area of finance and credit scoring. He has now taken his research and analytical skills into the business world and is currently employed by Marks & Spencer Financial Services in their Credit Development Department where he assists in the analysis and modelling relevant to their credit lending decisions.

Willem E.Saris is full professor in Methodology at the Social, Cultural and Political Sciences faculty of the University of Amsterdam. He has worked in the field of Structural equation models, especially with respect to models for evaluation of quality of survey questions. He has published papers and a book on this topic and has done substantive research in the field of Political decision-making and life satisfaction.

Judea Pearl is a Professor of Computer Science and Statistics at the University of California Los-Angeles, where he is the Director of the Cognitive Systems Laboratory. His areas of research include heuristic reasoning, probabilistic networks, causal discovery, and theories of actions and counterfactuals.

Professor Pearl is the author of 'Heuristics' (Addison-Wesley, 1984), 'Probabilistic Reasoning in Intelligent Systems' (Morgan Kaufmann, 1988) and 'Causality' (forthcoming, 1999). He has published close to 200 research articles, and is serving on the editorial boards of Artificial Intelligence, Annals of Mathematics and AI, and the Encyclopedia of AI.

Professor Pearl is a recipient of the RCA Laboratories Achievement Award (1965), NATO Senior Fellowship in Science (1975), UCLA 81st Faculty Research Lecturership (1996) and IJCAI Research Excellence Award (1999). He is a Fellow of the IEEE, a Founding Fellow of the American Association of Artificial Intelligence, and a member of the National Academy of Engineering.

John E. Cornell received his PhD in psychology from the University of Southern Mississippi, and he completed two years of postdoctoral training in behavioral statistics under the direction of Roger Kirk at Baylor University. He is an Associate Professor in the Department of Medicine at the University of Texas Health Science Center at San Antonio and he is the chief biostatistician for the Geriatric Research, Education and Clinical Center within the South Texas Veterans Health Care Center, Audie L. Murphy Division, in San Antonio, Texas. He also serves as the resident meta-analyst for the VA Cochrane Center at San Antonio. His primary research interests are in the design and analysis of studies involving longitudinal data. Within meta-analysis, he invests time in statistical models to assess and account for heterogeneity — including hierarchical Bayesian models — and in statistical models for combining information from correlated observations (multivariate, time-to-event, and repeated measures).

Cynthia D. Mulrow received her medical degree, residency training, and chief resident experience at the Baylor School of Medicine, Houston, Texas. She completed a General Medicine Fellowship as a Mellon Fellow at Duke University Medical Center, Durham, North Carolina. She was then awarded a Milbank Research Scholar award and received a MSc in Epidemiology from the London School of Hygiene and Tropical Medicine.

Dr Mulrow currently is the Chief of General Internal Medicine at the Audie Murphy Veterans Affairs (VA) Hospital, and a Professor of Medicine at the University of Texas Health Science Center at San Antonio. She is a graduating Senior Career Scientist in the VA Health Services Research and Development Program and is currently Director of the Agency for Health Care Policy and Research-sponsored San Antonio Evidence-based Practice Center and the Cochrane Center at San Antonio.

Dr Mulrow's special areas of interest are systematic reviews and guidelines, primary care, hypertension, and depression. She was co-editor of a special ten-part series and book on systematic reviews produced for the Annals of Internal Medicine. Currently, she is an associate editor for the Journal of General Internal Medicine for the 'From Research to Practice', an editorial board member of the British Medical Journal and ACP Journal Club, and a member of the US Preventive Services Task Force.

Chapter 1

Introduction

How this book was born... The idea for this book originates from a study on statistical expert systems that centered around the question why such systems are hardly ever used. One of the answers to that question was: the conceptual level at which such a system should function is *not* of a statistical nature but of a higher, more methodological one.

To put it in a more popular way: if you want a computer to act like a statistical expert, it should learn not only how to handle Statistics but also how to handle high-level methodological issues.

Of course, a computer is much more stupid than a human being and it requires very strict instructions as to what it has to do. On the other hand, if one can formulate rules for a computer, humans can certainly understand and apply them.

The research mentioned above forms only a slight pretense to write this book. To find out the fundamental rules that underlie the activity called 'scientific research' seems a worthwhile undertaking in itself.

Someone may argue that it is not so obvious that there is some underlying method to all empirical research. We think there is. And one of the aims of this book is to show there is. In the discussion at the end of the book we tried to dig out a general framework that encompasses the essential similarities and the obvious points of difference in the fields covered.

Thus, the point of departure of the book is the view that the Methodology of quantitative research is a discipline of its own, of which the basic notions, procedures and ways of reasoning can be precisely described.

About the title. By 'Research' we do not mean qualitative research, although we recognize that the approaches are closely related. Furthermore, the subtitle mentions a broad range of disciplines. One may ask whether it is too broad. Which fields of empirical research are *not* included? Essentially, none of the 'classical' empirical sciences is intended, like Physics or Astronomy.

To whom is the book addressed? Researchers involved in practical research tend not to make use of the methodological attainments of the last decade (which, by the way, are overwhelmingly many, as you will notice while reading through this book). This is not only because they are not aware of this progression but also since they perceive Statistics as a separate discipline with no obvious relationship to the solution of the problems they are facing. Since they approach their work in a practical way, their approach of statistical applications is practical, often resulting in a data analytical, technique-driven view.

It is easy to understand why the ongoing statistical discourse is of no direct consequence to the 'researcher in the field': It is often theoretical and not even methodological oriented.

This book may serve to bridge the gap between those two distinct points of view: the practical one directed towards doing practical research in a specific area like Experimental Psychology or Radiology. And, on the other hand, the point of view of the methodologist or the statistical advisor who is interested in responsive application of newly developed techniques.

For the researcher in the field, it describes the basic methods he or she can apply. Since the viewpoint is methodological, no deep statistical knowledge is required and, indeed, the contents are mostly closer to the actual research problems than a standard statistical text would be.

The book also provides a useful text for students of methodology and statistical consultation, since it describes in a unified way how to design and analyse research in various fields of the social and behavioral sciences and in Medicine. The student may learn what he can expect when he gives actual advice and which issues deserve special attention.

We would be very satisfied when the reader of this book would react somewhat like: "I got a better idea of what is available in this field, I learned something that I can use in my own research. If something is not in the book and I want to know more, I can consult the state-of-the-art references of the 'Further reading' sections the chapters provide".

Basic Idea. The informal title we used between us was: 'Methodology as a discipline' (In Dutch: 'Methodologie als vak'). This may seem amazing, considering the fact that we both call ourselves 'methodologists'. The reason was that when you call Methodology a discipline, it must be possible to indicate basic notions and ways of reasoning used by everyone who does research. When we tried to indicate what these basics might be, this turned out not to be straightforward at all.

We suspected there exist research paradigms, deeply anchored in the philosophy of science on which researchers subconsciously are oriented when they do their work. Therefore, our main question was

> *'What are the basic elements, the basic lines of reasoning and the paradigms underlying research methods?'*

The book is an attempt to discover these basic notions and to describe them. We do not deny that their development closely corresponds to, even coincides with developments in Applied Statistics and to research methods in all those specific areas of empirical research that use statistics during design and analysis.

We are well aware that in most disciplines a substantial literature on research methods exists that in some cases is rapidly expanding. But we also noticed that much less has been published on differences and similarities between the methods used in those application fields. The aim of the book is to do just that: to give the reader an impression of what methods are used in those fields and how they correspond or differ.

How did we proceed? As you can see, a considerable number of authors were asked to contribute a chapter to this book, each with his own views and expertise. This has pros and cons. Of course, it is downright unique to have contributions of people that have been working in their field for a long time and that have a lot of specialized knowledge, and on top of that can write comprehensively on the subject they love.

On the other hand, we ran the risk of getting a heterogeneous whole with contributions of different level and style. Also, from the point of view of the authors it is difficult to lay links to other subjects outside their field.

For the editors this resulted in a decision problem: to choose between readability on the one side and depth and personal style on the other.

We solved this in two ways. First of all, we discussed at length the contents and the structure of the book and decided how links between chapters should be established. We discussed this with the authors and tried to have them discuss the relevance of their field to the outside world, and, more particular, to the contents of the other chapters. In fact, the authors were supplied with some basic questions that could serve as a guideline during writing:

- *What are the basic notions and ways of reasoning in your field?* This, in a nutshell, is what the book is about.

- *What is your own position in your field?* This question was included since we thought it much more interesting to read about the specific themes an author is working on than about well-known principles that can be found in any textbook. It also took away the obligation for the authors to give a full account of their field in an insufficient number of pages.

- *Can other fields benefit from the findings in your field and are methods from other fields applicable in your own field?* This question was meant to offer the possibility to indicate links to and from other disciplines.

General structure. Apart from a historical introduction (Chapter 2) and one on meta-analysis at the end (Chapter 13), three parts may be discerned.

- In the first part *REPRESENTATION* is at the center. Various graphical representation methods are discussed in Chapter 3 and high level representations of data in Chapter 4 on meta data.

- The second part concerns *DESIGN* and describes a number of fundamental research designs: Experimental Design, Clinical Trials, Cross-sectional designs and Longitudinal Designs.

- The third part focuses on *MODELLING*. Like the part on Design, it has also four chapters, discussing respectively (a) Measurement models, (b) Graphical models, (c) Structural Equation models and, (d) Causal models.

Care has been taken to make the second part of the title '...in the Behavioural, Life, and Social Science' credible by taking examples from different subject matter fields. Of course, there is an end to this: most authors are familiar with one of those fields only and it would turn out unnatural to ask them to describe research outside their field of application. But most authors were quite able to illustrate the line of thought with a small sample from another field.

Are we satisfied with the result? The writing of the book went off unexpectedly well considering the many people that participated. Working on the project has been very rewarding both for the authors and the editors.

If we have in fact been able to point out what the foundations of this new, unified field called 'Research Methodology' should look like, remains to be seen. We undoubtedly produced a worthwhile overview over what is going on in various subfields. And this is the texture that the more fundamental stuff should be made of.

Acknowledgements. Special thanks go to David Hand who not only has been very stimulating during the whole project but who did carefully read all material and provided us with his sound and often very fundamental commentary.

All chapters have been reviewed by especially selected reviewers who, like the authors of the chapters, are specialists in the field covered by the chapter. Their comments have been invaluable. There is no way in which we could have done our job if the reviewers had not first commented on the abstracts and drafts, since it would have been virtually impossible to properly judge and value the importance of all the specific issues the chapters deal with. More importantly, not being an expert in the field, we would have been unable to spot which important topics had been left out.

Only the amount of space available and the emphasis the author of the respective chapter wanted to put on particular issues made us depart from reviewers' advice. Thus, the final responsibility for the contents of the chapters is the authors' and our own.

Amsterdam, May 31, 1999
Herman J. Adèr
Gideon J. Mellenbergh

Chapter 2

Some Remarks on the History of Method

Jaap van Heerden

It is interesting to note that a successful scientific achievement in the past is by historians of science appreciated for quite different reasons. Sometimes this appreciation differs considerably from the appreciation the scientist concerned originally wanted to attain. The reputable historian of psychology, E. C. Boring, cites Pascal's experiment in 1648 with Torricellian tubes at the Puy-de-Dôme as an early example of the use of control observations as a standard of comparison for the principal observations. "In 1648 the Torricellian vacuum was known to physics in general and to Pascal in particular. This is the vacuum formed at the upper closed end of a tube which first has been filled with mercury and then inverted with its lower open end in a disk of mercury (...). Pascal was of the opinion that the column is supported by the weight of the air that presses upon the mercury in the disk (he was right: the Torricellian tube is a barometer) and that the column should be shorter at higher altitudes where the weight of the atmosphere would be less" (Boring, 1954, p. 577). By measuring the height of the column at the foot of the Puy-de Dôme, at the top and at intermediate altitude one could collect evidence for the correctness of the hypothesis involved. It should, however, be noted that it was not Pascal's exclusive aim to prove that the Torricellian tube was a barometer. His aim was to prove in this way that the alleged horror vacui in nature does not exist. That was his principal purpose.

The horror vacui thesis stems from Aristotle. It says with so many words that nature does not tolerate and even abhors a vacuum. The actual Latin phrase was coined by Johannes Canonicus in the Thirteenth century. The thesis was still a matter in dispute in Pascal's days. In a letter to his brother-in-law, Monsieur Périer, he wrote "j'ai peine à croire que la nature, qui n'est point animée, ni sensible, soit susceptible d'horreur [du vide], puisque les passions présupposent une âme capable de les ressentir, et j'incline bien plus à imputer tous ces effets à la pesanteur et pression de l'air ..." (Pascal, 1954, 393) (I find it hard to believe that nature, not being animated nor sensitive, should be susceptible of horror [vacui], because passions presuppose a soul capable of feeling them and I rather incline to attribute all these effects to atmospheric weight and pressure). He then asked his brother-in-law, living in Clermont near the mountain Puy-de-Dôme, to perform an experiment, in the sense described above. It is Monsieur Périer who can take the credits for the well-balanced experimental design. He set up two tubes at the foot of the mountain, found that the columns were equal in length, left one at the base watched by an assistant in order to register possible changes during the time of the experiment, carried the disassembled other one to the top where it was put together. In fact, Périer ascertained in this way the indispensable control observations. "How important (...), How wise (...), How intelligent (...)", comments Boring (op.cit. p. 578). He obviously appreciated Périer's methodological sophistication. Pascal, on the other hand, explicated another methodological principle: an experiment should be decisive between two hypotheses. "vous voyez déjà sans doute, que cette expérience est decisive de la question, et que, s'il arrive que la hauteur du vif-argent soit moindre au haut qu'au bas de la montagne (...), il s'ensuivra nécessairement que la pesanteur et pression de l'air est la seule cause de cette suspension du vif-argent et non pas l'horreur du vide ..." (op.cit. p. 394) (You see for sure that this experience is decisive as to the question and that if it turns out that the height of mercury is less at the top of the mountain than at the foot (...) it necessarily follows that the atmospheric weight and pressure is the sole cause of that suspension of the mercury and not the horror vacui). In Pascal's opinion it was a crucial experiment. Dijksterhuis (1961) argues that Pascal's expectation as to the status of this experiment is unfounded, because other interpretations are not completely excluded and Copi (1953) has convincingly shown that the consequences of a crucial experiment can in principle always be undone by changing one of the auxiliary premises or by introducing an ad hoc hypothesis that explains the unfavourable outcome away. The term *crucial experiment* is not a useless one, however, says Copi "Within the framework of accepted scientific theory which we are not concerned to question, a hypothesis can be subjected to a crucial experiment. But there is nothing absolute about such procedure, for even well-accepted scientific theories tend to be challenged" (Copi, 1953, p. 425).

There is one other lesson to be learnt from Pascal. He attempted to decide this question by quantitative measurement in an experimental setting. The essence of quantitative measurement is the production of numbers as Kuhn (1977) says. These numbers indicate a position on a scale and the difference between the numbers record a difference in reality. But one of the significant functions of measurement is, according to Kuhn, "that measurement can be an immensely powerful weapon in the battle between two theories (...)" and: "In scientific practice the real information questions always involve the comparison of two theories with each other and with the world, not the comparison of a single theory with the world. In these three-way comparisons, measurement has a particular advantage." (op.cit. p. 211). That was exactly Pascal's strategy.

As said above, Boring appreciated this experiment for its methodological finesse, "195 years too soon for the experimenters to have read John Stuart Mill's Logic." (op.cit. p. 572). He obviously had in mind that Pascal and Périer are to be admired for this experimental design long before the methodological rules of experimentation were explicitly and systematically formulated. That was done by John Stuart Mill. As to experimental inquiry, his aim was to give a complete "enumeration of the means which mankind possess for exploring the laws of nature by specific observation and experience" (Mill, 1862, p. 436).

Laws of nature stipulate causal relations between phenomena or events. The inquiry into lawlike relations has a two-fold character: one either tries to find the cause of a given effect or the effects of a given cause. In fact mankind possesses four means or methods to find the causal connection between events. Mill's exposition has become classical and nowadays every textbook of Methodology gives a concise survey of his findings. Also Boring's article on the history and nature of control contains a fine summary. But we will turn to the original text because of the illuminating examples Mill provides each method with and because of the fact that the original text contains modifications of methods obviously of some historical importance, but rarely mentioned in contemporary overviews.

Mill denotes antecedents by large letters of the alphabet and consequences by small letters. "Suppose, for example, that A is tried along with B and C and that the effect is abc; and suppose that A is next tried with D and E, but without B and C and that the effect is ade. Then we may reason thus: b and c are not effects of A, for they are not produced by it in the second experiment, nor are d and e, for they were not produced in the first. Whatever is really the effect of A must have been produced in both instances; now this condition is fulfilled by no circumstance except a." Mill coins this way of experimentation the *method of Agreement*, essentially consisting in comparing different instances in which the phenomenon occurs. Under all possible circumstances A is followed by a. So we conclude that A in all probability is the cause of a. Mill's example is the production of soap (a). When we bring an alkaline substance into contact with an oil (A) under a variety of circumstances, the result is "a greasy and detersive or saponaceous substance" (op.cit. p. 426). The method of Agreement is never conclusive. Mill considered his second method, the *method of Difference*, as more potent. The procedure is: "If the effect of ABC is abc and the effect of BC is bc, it is evident that the effect of A is a" (op.cit. p. 428). It is evident that the joint methods of Agreement and Difference make a strong case. In a variety of circumstances A is always followed by a and in a variety of circumstances not-A is never followed by a. Mill noticed that both methods are methods of *elimination*. "The Method of Agreement stands on the ground that whatever can be eliminated, is not connected with the phenomenon by law. The method of Difference has for its foundation, that whatever cannot be eliminated, is connected with the phenomenon by law." The method of Difference can not always be employed as in history or astronomy. It requires *manipulation.* The method of Agreement is the justification of painstaking observation. Mill's third method is the *method of Residues.* If we know that the effect of ABC is abc and the effect of A is a and the effect of B is b, the method of Residues implies that "subtracting the sum of these effects from the total phenomenon, there remains c, which now, without any fresh experiment, we may know to be the effect of C" (op.cit. p. 436). This method is of special importance in cases where the effect of c cannot independently be established or measured, but only indirectly as part of the effect of another phenomenon.

I know of at least one example in the history of psychology wherein this method of Residues is applied. It concerns the subtraction method as developed by the Dutch physiologist F. C. Donders (1818–1889). It was widely used in the field of mental chronometry. How much time do mental processes take? One can register the reaction time of a subject, asked to press as quickly as possible a button as soon he sees the flickering of a light. One can refine the experiment by introducing three lights of a different colour with three corresponding buttons to be pressed as soon as one of the lights flashes. Now the reaction time is longer. By subtracting the first reaction time from the second one gets, approximately, the time it takes to discriminate and to make the proper movement. This is essentially Mill's method of Residues applied to measurement. By introducing one button only to be pressed at the flashing of one specific colour out of three or more, a task which takes less time than the previous one, because the subject has not to choose between different movements, one can now by subtracting the first reaction time from the last, find out how much time it takes to discriminate (see for more details Kolk, 1994).

A special difficulty is constituted by causes "which we can neither hinder from being present, nor contrive that they shall be present alone" (op.cit. p. 437). A good example of this difficulty is the relation between body and temperature or heat. "We are unable to exhaust any body of the whole of its heat. It is equally certain that no one ever perceived heat not emanating from a body. Being unable, then, to separate Body and Heat, we cannot effect such a variation of circumstances as the foregoing three methods require" (op.cit. p. 438). But there is still a resource, because we can study the effect of *modification* of the phenomena under consideration. We can try to change or modify or vary the impact of the alleged cause and effect. We can do so by careful observation as nature happens to bring about a change, or experimentally if the circumstances allow such an intervention. "If some modification in the antecedent A is always followed by a change in the consequent a, the other consequences b and c remaining the same, or vice versa, if every change in a is found to have been preceded by some modification in A, none being observable in any of the other antecedents, we may safely conclude that a is, wholly or in part, an effect traceable to A..." (op.cit. p. 439). In case of heat we can attain the conclusion that an increase or diminution of the temperature leads to an enlargement or contraction of the body. This fourth method is termed by Mill as the *method of Concomitant Variation.* But Concomitant Variation is in itself not sufficient to establish which of the two varying phenomena is the cause and which the effect, and it cannot be excluded either that the two modifications are due to a common cause. "The only way to solve the doubt would be (...) by endeavouring to ascertain whether we can produce the one set of variations by means of the other. In case of heat, for example, by increasing the temperature of a body we increase its bulk, but by increasing its bulk we do not increase its temperature; on the contrary, we generally diminish it: therefore heat is not an effect, but a cause, of increase of bulk" (op.cit. p. 442).

Reading Mill one is struck by the elegance of exposition and the clarity of thought. But his was not the last word in Methodology. He was even sometimes a bit naive in his practical strategy how to compare the experimental and control condition. Take for example how he conceived of an experiment in which one could establish the deadly effect of carbonic acid gas. "If a bird is taken from a cage, and instantly plunged into carbonic acid gas, the experimentalist may be fully assured (at all events after one or two repetitions) that no circumstance capable of causing suffocation has supervened in the interim, except the change from immersion in the atmosphere to immersion in carbonic acid gas" (op.cit. p. 431). The previous state is the control condition and Mill has to warn the would-be experimentalist to act "as rapidly as possible". It did not occur to him to form two groups of birds and to give one of them the required treatment. (If one accepts this sacrifice as justifiable for the sake of science. But that is another question.) In Methodology it took some time before the *Experimental Design* was developed and valued of drawing randomly a sample from the population and assigning randomly subjects to the experimental and control conditions. That in itself was a great innovation (See Dehue, 1997). The random assignment of subjects to conditions meant, properly understood, an optimal realization that individual characteristics of subjects were equally distributed over the experimental and control group. The same could be realized by matching. This kind of experimental design was in psychology not realized until around 1900, partly because mainstream psychology was primarily interested in individual achievements of well-trained subjects, reporting by introspection on the performance of mental tasks, and partly because the statistical means were not yet fully developed to properly deal with group-differences.

Aristotle adhered to the rule that everything goes in sevens. This rule had to be considered as self-evident. Moreover, it was a useful device for analysing scientific problems. It revealed, for instance, that man has seven ages, each seven years long. This fundamental insight made it possible to deduce and prove all kinds of interesting theorems.

One could say that this septuple or hebdominal rule, although essentially dogmatic, was an early attempt to use an *algorithm* in matters of science.

In a primitive way it involves the assignment of numbers to phenomena, and the application of simple arithmetics warrants a fruitful continuation of empirical research. Aristotle's approach is a methodological innovation that deserves the requisite attention in the history of ideas but at the same time it shows that not all methodological innovations contribute to the advancement of science. It refutes the myth that the enhancement of scientific knowledge exclusively depends on the developments of methods. What the growth of knowledge requires is an ***independent analysis*** and ***justification of scientific methods*** (as Mill did), free from metaphysical assumptions as to the dominant place of one number or another. Bear in mind that the same metaphysical assumption about the pre-eminence of the number seven prevents regarding Aristotle's rule as a primitive form of measurement. It is impossible to see it as such, not only because of the queer consequence that all measures would yield the number seven as a result, but also because of the fact that the outcome of any measurement whatsoever precedes the factual process of measuring. The outcome is fixed and reality has to obey or to be accommodated. The best one can say of Aristotle's rule is that it is a heuristic advice: it may be profitable to view any phenomenon as consisting of seven parts. But that is more a form of modelling than a form of measuring.

Beginning with classification, i.e. ordering by comparison, Science step by step developed quantitative measuring in the sense that each item falling under a certain concept, occupies one and only one place on a numerical scale although more than one item could occupy the same place.

Francis Galton (1822–1911) has to be remembered for several outstanding contributions to the methodology of the life sciences.

In the first place he observed that eminent mental ability frequently goes by descent and seems to run in certain families. By studying family-histories he found that the proportion of highly talented relatives exceeded the proportion to be expected by chance. Being a cousin of Charles Darwin and probably strongly influenced by his cousin's 'Origin of Species', Galton became a strong believer in the heredity of intelligence, talent and character. Mental characteristics could in principle to the same extent be measured as physical ones. Galton adopted Quetelet's method of fitting normal curves to data. As Stigler (1985) says: "He was fascinated by the appearance of the 'very curious theoretical law of deviation from an average' in so many different cases, for heights and chest measurements and even such measures of talent as examination scores. Following Quetelet, he proposed that the conformity of the data to this characteristic curve was to be a sort of test of the appropriateness of classifying the data together in one group; or rather the non-appearance of this curve was indicative that the data should not be treated together".

In a way one could also regard Galton as the inventor of the mental test. He took measurements for association, by registering the reaction times needed to respond to a target word. But more important still is his invention of the measure of *correlation*, for which he gave a calculation instruction. He applied his method to variations in measurements of diverse parts of the human body: tallness and long head length, for instance, go together. The upshot of his scientific endeavours was a political manifesto. He wrote "What an extraordinary effect might be produced on our race, if its object was to unite in marriage those who possessed the finest and most suitable natures, mental, moral and physical" (Galton, 1865). He became the founder of eugenics and the eugenic movement. In this he was followed by his disciple Karl Pearson (1857–1936), who also improved on his calculation of the correlation coefficient, the well-known Pearson's product moment estimate of the correlation coefficient. Given the deplorable state of the British Empire, Pearson formulates the remedy: "getting the intellectual section of our nation to realise that intelligence can be aided and be trained, but no training can *create* it. You must breed it, that is the broad result for statecraft which flows from the equality in inheritance of the psychical and physical characters in man" (Pearson, 1903). The debate on this issue is still with us.

Chapter 3

Graphical Representation of Methodological Concepts

Herman J. Adèr

What may be the reason for representing methodological concepts in a formal way? It is not difficult to think of situations in which a reliable and easy-to-use formal representation could be beneficial. For instance, since *Planning* largely determines the research design and thus the data analysis and the interpretation of the results, an unambiguous representation of the research plan could help to avoid design flaws.

As an example, think of Computer Assisted Interviewing (CAI). Assume we were able to represent the order in which the questions have to be put to the interviewee as a decision tree. The very activity of drawing such a tree will uncover any omission.

Another example is the design of a protocol for a Clinical Trial in Medicine. Here also, a formal description could help to spot ambiguous points in the whole procedure during the trial and specify it in complete detail.

Now an obvious next question is: *What exactly should be imagined by 'Representation'?*

In Artificial Intelligence the term 'Knowledge Representation' is used to indicate formal ways to lay down knowledge of experts in a specific domain. Starting from this representation more or less intelligent computer software can be built that mimics the approaches of these experts. Many methods to describe high level concepts have been proposed in this field, several of which will be considered in the next sections.

. Some notions are too frequently used in Artificial Intelligence not to briefly mention them here. The set of entities composing the knowledge of concern is called ***universe of discourse***. A distinction is made between ***declarative*** and ***procedural*** knowledge. The first kind refers to the plain definition of the objects in the universe and the relations between them. The second kind refers to the actions that may be taken upon the defined objects

and their relations. *Strategic* knowledge is a special kind of procedural knowledge, concerning the way to use declarative and procedural knowledge to attain some predefined *goal*. An example is the way the data analyst goes about in the series of analyses he or she performs (Oldford, 1990). Genesereth & Nilsson (1987, Chapter 12) explain the theoretical side of the representation of strategies in their chapter on Planning.
Strategic knowledge will be discussed in Section 3.5.1.

With the representation of methodological knowledge two concepts play a key role: *Context* and *Time*. Of course, the notion of Context is important in many other fields. What makes this case a special one is that we can indicate what the context consists of. Apart from ideas and notions from applied statistics, many elements of the context originate from the research field ('the subject matter field') in which the basic research question was posed.

As to the notion of Time, for many study designs temporal issues are essential, be it to study variability over time (in reproducibility studies and studies on responsiveness) or to study development of a phenomenon over time. This may be on a per person basis as occurs in a repeated measurement design or in a more epidemiological sense in a cohort study. A host of statistical methods is available to appropriately handle the resulting data: The analysis of repeated measurements, Trend analysis, Time series analysis, Longitudinal analysis, and the analysis of survival data are all well-studied statistical techniques, specially devised to analyse temporal data.

Therefore, both Context and Time are notions that should be expressible in any representation system. In this chapter, this will be touched upon in Section 3.3 on Time and Section 3.5 on Context.

Although the chapter discusses a formal subject, the approach will be rather informal. The reason is that I will concentrate on *graphical* representation methods that invite a casual discussion. I even restrict myself to a particular kind of diagram, called *planar graph*. This is a graph placed in the two-dimensional plane that has as its building stones *nodes* and *edges* between nodes.

In Section 3.1 different ways to graphically represent statistical and methodological concepts will be staged. We will try to find out what their respective merits are in provisions to the express methodological notions and activities. In Section 3.1.1 some general requirements will be formulated that seem essential. In Section 3.2 basic ideas of so-called *Functional notation* (Adèr, 1995, Chapter 5 and 6) will be given. In the subsequent sections, this notation is used for the representation of temporal notions (in Section 3.3), reasoning (in Section 3.4), strategy (in Section 3.5.1) and context (in Section 3.5.2).

This introduction ends with a word of warning: Diagrams are often supposed to be self-explanatory and easy-to-comprehend while implicitly it is assumed that all elements shown as well as their interrelations are well-defined. But very often the formal correspondence between a diagram and its meaning is not trivial at all. One of the aims of this chapter is to clarify where and when graphical representation methods can be usefully applied without leaving much room for ambiguity.

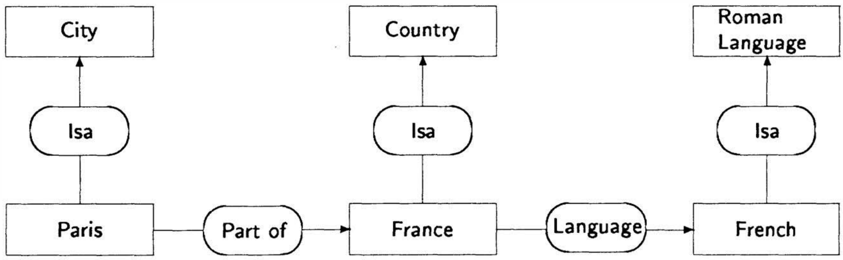

Figure 3.1: Example of a semantic net.

3.1 Common approaches in various fields

In artificial intelligence, a variety of graphical representation styles of conceptual data models have been proposed, many of them derived from so-called *semantic nets* which will be discussed hereafter. Other approaches have been developed in the realm of social science research: I will comment on graphical methods to represent experimental design, on structural equation models and graphical models. At the end of this section the usual graphical representation of Artificial Neural Networks are considered.

Semantic Nets, Conceptual Graphs and Concept Mapping. Figure 3.1 gives an example of a semantic net[1]. The boxes (`City`, `Paris`, `French ...`) contain Concepts, while the arrows are labelled to indicate relationships between concepts: `Isa`, `Part of` and so on.

The idea of semantic nets go back to the work by Peirce (See Peirce (1933)). Being a logician, Peirce constructed a graphical notation for predicate calculus. Other graphical systems for general knowledge representation have developed from this. Quillian (1968) was first to formulate a symbolism in connection to cognitive models for memory organization. This forms the basis of later knowledge representation methods. Other areas of application are language research (for a recent review, see Willems (1993)).

Although the representation differs slightly, *conceptual graphs* are similar to semantic nets. Sowa (1984) formulated the syntax and semantics that can be used to build such graphs. In the conceptual graph representation all important facts are grouped around nodes, so that a two dimensional

[1] I preferably adhere to the way graphs are usually drawn in the field they are taken from. If no particular conventions seem to exist, I will use *Functional notation* as described in Adèr (1995). A succinct description of this formalism will be given in Section 3.2.

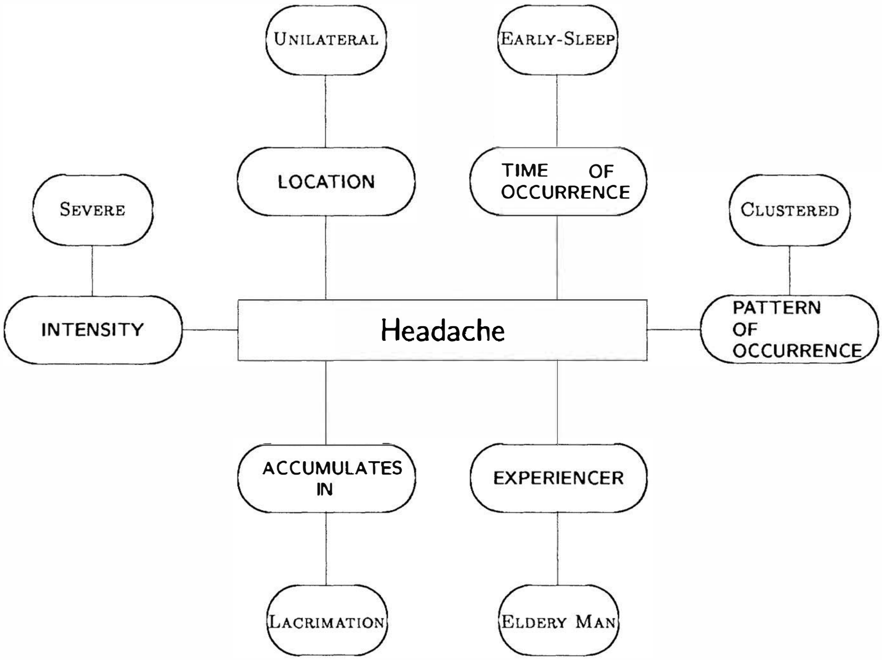

Figure 3.2: Conceptual graph, describing headache complaints. Source: Clancey (1985).

structure arises. There are two basic elements: *notions* and *relations* between notions, which are both labelled. In Figure 3.2 the notion 'Headache' is depicted in the middle of a net of related predicates ('severe', 'unilateral' 'early-sleep' and so forth). Relations have uppercase labels like 'intensity', 'location', 'time of occurrence', respectively). In the figure, relations are indicated by undirectional edges with a label inside an oval, while the predicates are indicated by ovals with small caps as a type font. Note that what is depicted here is the data of a single patient. Substituting other values for the predicates gives the representation of another patient suffering from the same disease. Putting variables over a set of predicates in place of the present values would extend the present representation to indicate a group of patients and would make it *generic*.

Conceptual graphs may well be used to describe subject matter concepts previous to the formulation of an empirical research design. Thus, a generic version of Figure 3.2 could be interpreted as a formal description of the complex structure of headache complaints previous to the formulation of the design of a clinical trial to investigate the disease[2]. Although this graphical formalism is effective as a means to depict complex phenom-

[2]Clancey (1985) introduces conceptual graphs as a formal specification after which to construct diagnostic computer software.

ena, some aspects are difficult to express. In particular, logic operations like negation and disjunction and procedural knowledge can not easily be incorporated.

As stated in the introduction, any representation of methodological knowledge, be it graphical or not, should have provisions to represent objects placed in time. Conceptual graphs have only imperfect provisions for this, since time is handled like any other attribute and thus it is difficult to indicate how more complicated structures behave in time.

Authors like Novak (1990) have extended the idea to a related representation called *Concept maps*. It was used to describe the knowledge context of learning. In this area a lot of effort has been invested in developing computer programs with which these kinds of graphs can easily be constructed (See Lanzing, 1996, for an evaluation of existing software). In Figure 3.3, a screen dump is given of one of these programs: again a concept (the starlike figure) is surrounded by a net of connected notions. Note that in contrast to the conceptual graph of Figure 3.2, directional arrows are used instead of links.

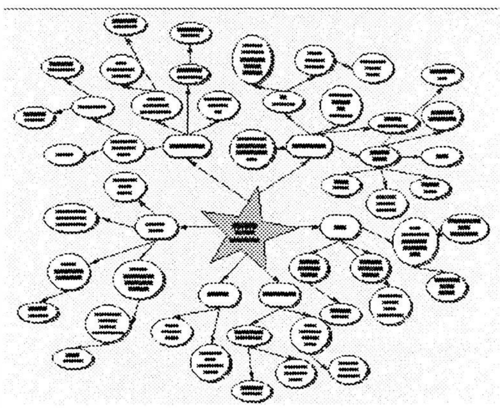

Figure 3.3: Screen dump of Inspiration©, a program to construct concept maps.

Experimental Design. Figure 3.4 shows the way Cook & Campbell (1979) indicate the designs of (quasi-)experiments. (a) indicates the basic experiment: Observation followed by treatment, followed by another observation. In (b) a control group is present, on which the same observations are done, but no treatment is given. The dashed line indicates that the experimental groups are not randomly formed. In picture (c) the precondition is observed in a different (but comparable) population. Note that the figure differs from other figures in this section in that it is not built from nodes with interconnecting edges.

A modern graphical representation of an experiment is given in Figure 3.5. This representation gives an impression of the general structure of

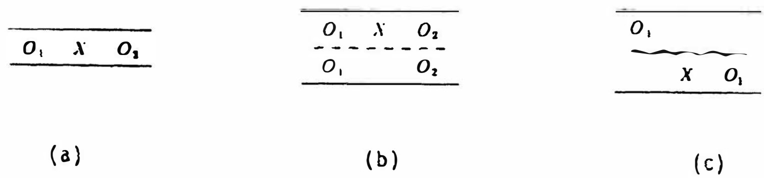

Figure 3.4: Notation to indicate quasi-experimental designs.
Legend: O: Observation, X: Treatment. (a) Observation, treatment, observation; (b) Non-randomized design with a control group; (c) Design in which the precondition is observed in a different (but comparable) population. Source: Cook & Campbell (1979).

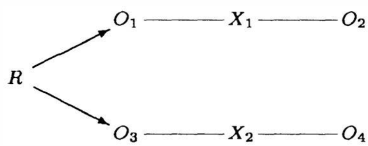

Figure 3.5: Experimental design: before–after two-group design. R: Randomly assigned subjects; O_i: Observation; X_i: Treatment (original caption).

a randomized clinical trial (See Judd, Smith, & Kidder (1991) for several alternative designs.) After randomization, subjects are assigned to either one of two treatment arms.

Both Figure 3.4 and 3.5 are ambiguous (although they may be effective if benevolently interpreted). In particular, in Figure 3.4, information on the structure of the experiment is intermingled with information on experimental units: in Figure 3.4 dashed and wavy lines indicate population characteristics, while the figure itself describes the temporal arrangement of the observations.

In Figure 3.5, R represents a group of subjects that have been already randomly assigned while O_i and X_i are activities *one* patient is subjected to. Note that this makes the meaning of the arrows unclear: The group of randomly assigned subjects cannot possibly *produce* observations O_1 and O_3. The figure would be more comprehensible if we were allowed to read R as '*one* subject to be randomly assigned'. R and the two arrows originating from R could then be interpreted as representing the randomization procedure and the subsequent assignment to either of two arms. In this case we are left with the interpretation of the links between observation and treatment and vice versa: we would have been happier when arrows had

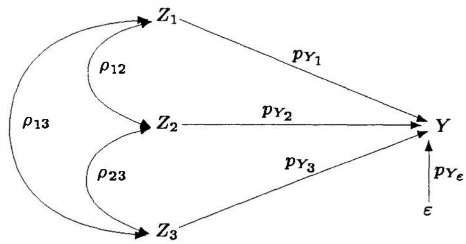

Figure 3.6: Example of a path diagram of a regression model. Y: dependent variable; z_i: independent variables (predictors); ρ_{ij}: correlation coefficients; $p_{Y_i}, p_{Y_\varepsilon}$: path coefficients.

been used here, since now the reading order in the figure implicitly indicate the arrangement in time. Admittedly, this is also indicated by the indices of observations and treatments: odd indices indicate the first administration while even ones indicate the second administration. But when we compare this to Figure 3.4 in which comparable observations have the same index, it is no longer clear whether O_1 and O_3 indicate the same kind of observation (this also holds for O_2 and O_4).

Finally, to end this niggling discussion, note that the nodes in this figure do not refer to objects but rather to activities, R being the act of randomization, O that of observing and X of giving treatment. Later on, in Section 3.2, a detailed representation of this design is given.

Path diagrams and Structural Modelling. Wright (1934) introduced so-called *path diagrams*, to visualize regression models. In Figure 3.6 an example is given. The objects that occur in this particular figure are variables Z_i and Y and the error term ε. The ρ_{ij}s labelling the bidirectional arrows are product-moment correlation coefficients indicating the strength of association between covariates Z_i. The path coefficients p_{Y_i} and p_{Y_ε} indicate the strength of the influences of the covariates and unsystematic error on the dependent. Thus, if we have to formulate what meaning should be attached to the arrows in the figure, we could read the bidirectional arrows as 'is associated with' and the directional arrows as 'is influencing'.

Wright's work led to the introduction of 'Structural Equation Modelling' (SEM by abbreviation) in which similar graphical representations are used (See Chapter 11 and 12 for a more extensive account of the historical background of path diagrams and Structural Equation Modelling).

SEM offers a wide variety of modelling possibilities. Confirmatory factor analysis, regression analysis and perhaps most importantly, models with latent parameters, can all easily be represented. It is a way of analysing in which both methodological and statistical considerations play an important part. As such, this field has much potential when it comes to the formulation of the basics of research methodology. Initial interest in the technique was raised when some authors tried to assess causal relationships and used

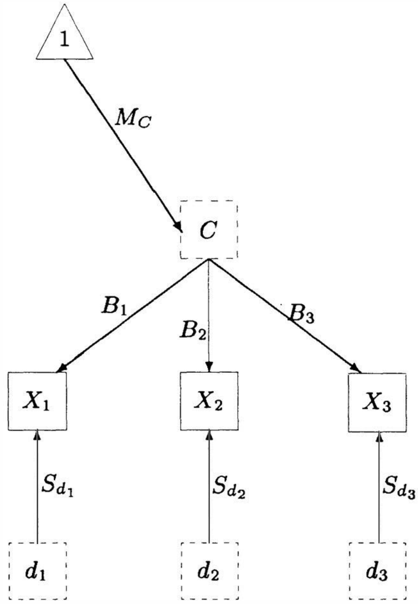

Figure 3.7: Structure diagram for a CURVE factor model. 1: constant; C: latent factor, X_1, X_2 and X_3: manifest variables, d_1, d_2 and d_3: measurements at time $t = 1, 2, 3$. B_i, S_{d_i}: parameters to be estimated. Source: McArdle & Epstein (1987).

structural equation modelling for it. This direction is now usually called *causal modelling*. Much more on the subject is discussed in Chapter 12.

As argued before, it is important that a representation method allows to express temporal notions. The following example, taken from McArdle & Epstein (1987), demonstrates the usual approach to model temporal aspects in SEM. It is an application of structural models in the field of child development research. Figure 3.7 gives what they call the 'CURVE' model that assumes a latent factor C, with manifest variables X_1, X_2 and X_3, measured by d_1, d_2 and d_3 on three occasions ($t = 1, 2, 3$).

Note the way in which time-related aspects are modelled: by duplicating objects and indexing for time. The notation is not different from a model with three observed variables without repeated observations. In Section 3.3 an alternative representation method will be given in which temporal aspects can be indicated more clearly.

Graphical modelling. Whittaker (1990) discusses statistical models that use a graphical representation of the (in)dependence structure of the variables. This will also be the subject of Chapter 10. A variety of statistical models can be represented this way including Markov chains. Structural equation models may be interpreted as a special case.

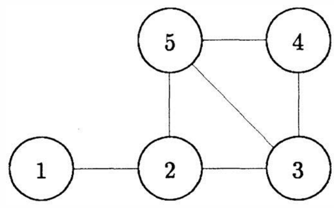

Figure 3.8: Dependence diagram of the logit model: $\log p_{12345} = u_\phi + u_1 + u_2 + u_3 + u_4 + u_5 + u_{12} + u_{23} + u_{25} + u_{34} + u_{35} + u_{45} + u_{235} + u_{345}$.

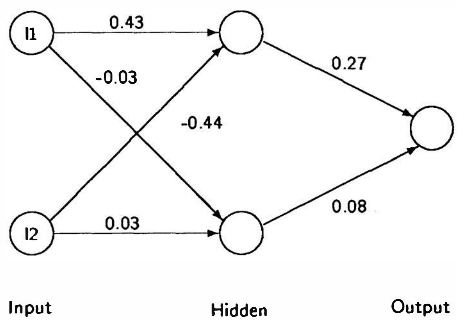

Figure 3.9: Neural network implementing the XOR-function.

In Figure 3.8 an example is given of a logit model with five main effects. The objects are variables indicated by a number placed in a circle. The links (undirectional arrows) between nodes should be interpreted as 'are interdependent'. Note that the term u_{24} is not present in the model, according to the absence of a link from node 2 to node 4. It is not possible to distinguish in this way between the model shown and a model from which third order terms are left out.

Artificial Neural Networks (ANNs). Since applications of neural networks in statistics are not uncommon any longer, we consider their graphical representation here, too. Hassoun (1995), who gives a thorough mathematical treatment of this kind of computational structure, characterizes artificial neural networks in the following way: ... *"parallel computational models comprised of densely interconnected adaptive processing units."* Applications of neural networks in Statistics and pros and cons thereof are discussed in Ripley (1993). Adèr & Bramsen (1998) discusses connections between neural networks and structural equation models.

Usually, a distinction is made between *feed-forward* and *interactive* networks. Both kinds can be represented as a directed graph. An interactive network may have bidirectional arrows, while in a feed-forward network only arrows in one direction occur and cycles are not allowed. Figure 3.9 gives an example in the usual graphical representation. Note that a feedforward network has at least an input and an output layer. Often, there are also one or more hidden layers. The nodes of the same layer are not interconnected.

In many cases, three phases can be discerned in the use of the network: a *training* phase in which network weights are calculated using a training set consisting of input patterns and corresponding target patterns that should be delivered as output; a *testing* or *generalization* phase in which new patterns including the desired output are processed without adapting the network weights but calculating some error function to assess the adequacy of the already trained network; and an *application* phase in which the optimally trained network functions as a device to calculate output patterns from new input patterns without providing target patterns. In Figure 3.9, the arrows of the network have weights attached that have been calculated during training.

During application, information flows through the network in 'waves': when an input pattern comes in, the input layer is activated which passes on information to the hidden layer. When all processing in the input nodes is finished, the hidden layer takes over and information passes to the output layer which produces the desired output pattern.

In Figure 3.9 a network that functions as the XOR-function (exclusive OR) is shown. This network has one hidden layer. The arrows are validated by weights. It is easy to see that the network functions as a XOR operator: for instance, when the input is $(I_1, I_2) = (1, 1)$, the upper hidden node receives $0.43 + (-0.44) = -0.01 < 0$ and therefore does not pass on information. The input to the lower hidden node is $-0.03 + 0.03 = 0$. Since both hidden nodes do not produce output, the output node is not activated and the network yields 0.

The term 'parallel' in Hassoun's formulation refers to the possibility that nodes of one layer process (or are processed) in parallel. Parallel execution is possible since in all phases layers are processed sequentially and nodes of one layer are not interconnected.

It may be clear from the above description that function and meaning of the nodes and arrows in the graphical representation of a neural network differ between phases. Generally speaking, one may say that:

- each node is able to perform a distinct action which may be different between layers or modules. For instance, a node in the hidden layer should be able to sum the weights of the incoming edges multiplying them by the value of the corresponding input nodes (typically: 0 or 1).

  During training, the computed output pattern as a whole has to be compared to the target pattern and weights of the network have to be adapted to decrease the value of the error function.

- The arrows of the graphical network representation are unlabelled. They have, however, an attached weight. During training, some operation on the arrow sets between layers should be available to minimalize this error function, The operation changes the weights of the arrows accordingly.

The meaning of both nodes and arrows of a neural network cannot be understood by reading the graph: one needs to know the phase the network is in and even then the meaning of the nodes may differs from layer to layer and between arrows. In the training and evaluation phase, operations occur that are applied on layers or even on the whole network.

3.1.1 What requirements should an ideal graphical representation fulfill?

As we saw in the previous section the basic elements of most graphical representations are objects and edges between objects. Sometimes these objects are indicated by letters in a box or a circle (as in the case of graphical modelling), sometimes only letters are used. As we saw in the discussion of Figure 3.5, sometimes nodes do not indicate static entities (objects) but *activities* (like 'randomize' or 'observe'). Since directed edges (arrows) are commonly used to indicate activities, it follows that the distinction between objects and arrows is not strict. Several kinds of edges are used: unidirectional, bidirectional and undirectional. Relations are usually represented by undirectional edges (links).

When figures are ambiguous, this is often caused by the *meaning* the reader has to attach to the edges. Sometimes they should simply be understood to mean 'produces', 'go to the next step' or 'pass on information', but sometimes, completely different interpretations are needed like 'influences', 'is dependent on' or in the training phase of a neural network, something like 'change the attached weights and pass on the result'. Usually, this meaning is not explicitly given with the figure and, for some dark reason, assumed self evident.

In Adèr (1995, Chapter 5) requirements are formulated to make any formalism to express methodological concepts more effective. Although the approach there is more general, it applies to graphical representations, too. We mention these requirements here without elaborating much on the rationale behind them: Several of them do not require much explanation, others will be highlighted in the rest of the chapter.

Recognizability. A representation should look familiar to those who use it. In particular, it should not deviate to much from existing formalisms.

Rough descriptions. A representation should stay meaningful even if the user doesn't care to specify the notions in full detail. In this way, rough indications can be given that are eventually completely specified at a later stage.

Handling complex pieces of information. It should be possible to label pieces of information of high complexity without requiring to completely specify the complex internal structure.