Federica Russo

Methodos Series 5

Causality and Causal Modelling in the Social Sciences

Measuring Variations

Springer

Causality and Causal Modelling in the Social Sciences

METHODOS SERIES

VOLUME 5

This Book Series is devoted to examining and solving the major methodological problems social sciences are facing. Take for example the gap between empirical and theoretical research, the explanatory power of models, the relevance of multilevel analysis, the weakness of cumulative knowledge, the role of ordinary knowledge in the research process, or the place which should be reserved to "time, change and history" when explaining social facts. These problems are well known and yet they are seldom treated in depth in scientific literature because of their general nature.
So that these problems may be examined and solutions found, the series prompts and fosters the setting-up of international multidisciplinary research teams, and it is work by these teams that appears in the Book Series. The series can also host books produced by a single author which follow the same objectives. Proposals for manuscripts and plans for collective books will be carefully examined.
The epistemological scope of these methodological problems is obvious and resorting to Philosophy of Science becomes a necessity. The main objective of the Series remains however the methodological solutions that can be applied to the problems in hand. Therefore the books of the Series are closely connected to the research practices.

For other titles published in this series, go to
http://www.springer.com/6279

Federica Russo

Causality and Causal Modelling in the Social Sciences

Measuring Variations

Springer

Dr. Federica Russo
Chargé de Recherche du FSR-FNRS
Université Catholique de Louvain
Institut Supérieur de Philosophie
Place Mercier 14
1348 Louvain
Belgium
federica.russo@uclouvain.be

Honorary Research Fellow
University of Kent
Cornwallis Building
Canterbury
KENT CT2 7NF
United Kingdom
f.russo@kent.ac.uk

ISBN 978-1-4020-8816-2 e-ISBN 978-1-4020-8817-9

Library of Congress Control Number: 2008931992

Cover illustration:

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

To my parents Paola and Sergio

Acknowledgements

The story of this book began some five years ago, when I was a Ph.D. student at the University of Louvain. It then continued back and forth between the UK and Belgium: at the Centre for Philosophy of Natural and Social Science (London School of Economics), in Louvain again, then in Kent, and finally back in Louvain where I am now. All those places have been ideal and stimulating environments. But places are not good or bad as such—what makes them enjoyable or tedious is the people you meet there. Philosophers, contrary to a widespread stereotype, are not (not nowadays, at least) lonely and isolated persons trying to build their own philosophical system, sunk in an armchair and locked in their room. If they are, I am certainly not of that kind (or perhaps I am not a philosopher?). Indeed, I have been lucky enough to benefit from rousing work environments in which to *do* philosophy—that is debating, discussing, arguing about philosophical problems—and thus to achieve what I hope are good results for social scientists and for philosophers.

The main outcome of this work is what I call the rationale of variation: quantitative causal analysis establishes causal relations by measuring variations, not by establishing regular sequences of events. I have worked hard to build empirical, philosophical, and methodological arguments to support this view. Needless to say, it is not for me to say whether and to what extent I have been successful in challenging the dominant Humean paradigm. Yet several people, notably Robert Franck and Daniel Courgeau, encouraged me to believe that this idea was worth a book—what is more, a book in the *Methodos Series*. Together we met in Tours in July 2005 and discussed at great length the structure and the contents of the book. I am extremely grateful to them for all the encouragement, advice, and feedback they gave me over the last few years. It is for me an honour and a privilege to present my work within this Series. I had hoped for a long time to publish my ideas in the *Methodos Series* because it is a long-term project that aims to bridge the gap between methodology of science and philosophy. There is no doubt this Series is a hub for a fruitful dialogue between philosophy and the sciences, and I am now an *active* part of it. That gives joy and satisfaction, and motivation to carry on the academic adventure.

Many other people, over the years, have helped me to clarify and strengthen the arguments presented herein. In particular, I would like to mention Michel Mouchart,

for his simple and yet deep questions—nothing is obvious to his eyes; for him, one question always hides at least three more. Jon Williamson, for his philosophical rigour and for always making the constructive objection you haven't thought about. Guillaume Wunsch, for his wise comments as a practising scientist as well as a philosophically minded methodologist. Over the last several years, their support has not just been academic but also personal. I wholeheartedly thank them for being some kind of (voluntary or involuntary) guardian angels. Likewise, I wish to thank Jossi Berkovitz, Julian Reiss and Jean Bricmont, for their sympathy and intellectual spurs. My gratitude also goes to Daniel Little, who kindly agreed to write a preface to the book. Finally, going back in time and space, when and where the story of this book began, I wish to thank Michel Ghins, for having pushed me into the impervious—and yet exciting—academic jungle from the very start.

At Springer, my gratitude goes to Evelien Bakker and Bernadette Deelen, for their patience in waiting for the final manuscript and for all their kind assistance at each stage, from the book proposal to the final touches with regard to the editing. I would also like to thank two anonymous referees, whose comments on the draft have been particularly helpful.

Various topics discussed in the book have been presented at conferences and seminars. Remarks, criticisms, and suggestions have all been remarkably valuable to me. I wish to thank all those who attended those events. It would be impossible to mention them all here, but particularly good feedback came from the summer schools in Konstanz, the seminars of the Evidence Project, the conferences at the IUC in Dubrovnik, various events at the LSE, Kent, and Louvain. Of course, any mistakes, imprecision, or inaccuracies remain mine.

Having finished the book, it is certainly thanks to the financial support I received from a number of institutions (FSR-UCLouvain in 2002–2005, British Academy and FSR-UCLouvain in 2006, University of Kent and FSR-UCLouvain in 2006–2007, FSR-FNRS from October 2007).

I cannot neglect to remember the many friends in Louvain, London, Brussels, Canterbury—some of them now spread over four continents—who constantly backed my perseverance, rightly opposed me when this perseverance was turning into being a workaholic, and cheered me with a good laugh. Last, but not least, I would like to thank my family. They are always with me and without them I simply wouldn't be where I am.

Contents

Preface xiii

Introduction 1
1 Scope of the Book and Methodology 1
2 Structure of the Book 3
3 Philosophical Issues at the Back of One's Mind 5
4 Philosophy at the Service of Social Research 9
5 Open Problems: Causal Realism, Objectivity, and Social Ontology 11

1 What Do Social Scientists Do? 15
1.1 Different Causal Claims? 16
1.2 Smoking and Lung Cancer 20
1.3 Mother's Education and Child Survival 21
1.4 Health and Wealth 23
1.5 Farmers' Migration 24
1.6 Job Satisfaction 25
1.7 Methodological and Epistemological Morals 25
1.7.1 Statistical Causality in the Social Sciences 26
1.7.2 Causal Claims in Causal Contexts 30
1.7.3 Testing Variations 31
1.7.4 Modelling Mechanisms 32

2 Probabilistic Approaches 35
2.1 Philosophical Accounts: Good and Suppes 36
2.1.1 Good (1960–1961) 36
2.1.2 Suppes (1970) 40
2.1.3 Good and Suppes: Common Features 43
2.2 Probabilistic Theories: Traditional Criticisms 44
2.2.1 Probabilistic Theories vs. Deterministic Theories 44
2.2.2 Negative Causes and Twofold Causality 46
2.3 Bringing Causal Theory to Maturity 50

3 Methodology of Causal Modelling 53
3.1 Methods and Assumptions of Causal Modelling 54
3.1.1 Path Models and Causal Diagrams 54
3.1.2 Covariance Structure Models 60
3.1.3 Granger-Causality 61
3.1.4 Rubin's Model 63
3.1.5 Multilevel Analysis 64
3.1.6 Contingency Tables 67
3.2 Hypothetico-Deductive Methodology 70
3.3 Difficulties and Weaknesses of Causal Modelling 75
3.3.1 Background Knowledge and the Formulation of Causal Hypotheses 76
3.3.2 Theorising or Modelling? 77
3.3.3 Lack of Theory 78
3.3.4 Establishing Social Laws 79
3.3.5 Concepts and Indicators 80
3.3.6 Causal Directionality 83
3.3.7 Searching Causes of Effects or Effects of Causes? 83
3.3.8 Underdetermination 84
3.3.9 Complexity of Causal Structures 85
3.3.10 Interpretation of Tests 86
3.3.11 Validity 87
3.3.12 Generalisability 90

4 Epistemology of Causal Modelling 91
4.1 The Rationale of Causality: Measuring Variations 92
4.1.1 Methodological Arguments 94
4.1.2 Foundations 101
4.1.3 Arguments from the Philosophical Literature 105
4.1.4 Objections, or Further Evidence for Variation 107
4.2 Varieties of Variations 114
4.3 What Guarantees the Causal Interpretation? 116
4.3.1 Associational Models 117
4.3.2 Causal Models 119

5 Methodological Consequences: Objective Bayesianism 133
5.1 Probabilistic Causal Inferences 134
5.2 Interpretations of Probability 135
5.2.1 Classical and Logical Interpretation 135
5.2.2 Physical Interpretation: Frequentist and Propensity 136
5.2.3 Bayesian Interpretations 137
5.3 The Case for Frequency-Driven Epistemic Probabilities 139
5.3.1 Probability as a Twofold Concept 140
5.3.2 Frequency-Driven Epistemic Probabilities 141
5.3.3 Credence-Driven Physical Probabilities 145

5.3.4 Objections . . . 146
5.4 Objective Bayesianism in Causal Modelling . . . 150
5.4.1 Empirically-Based vs. Objective Bayesianism . . . 150
5.4.2 Design and Interpretation of Tests . . . 151
5.4.3 Guide for Action . . . 153

6 Methodological Consequences: Mechanisms and Levels of Causation . . . 155
6.1 Mechanisms . . . 156
6.1.1 Modelling Mechanisms . . . 156
6.1.2 Mixed Mechanisms . . . 163
6.1.3 Explaining Through Mechanisms . . . 165
6.1.4 Modelling Causal Mechanisms vs. Modelling Decision-Making Processes . . . 171
6.2 Levels of Causation . . . 172
6.2.1 Twofold Causality . . . 173
6.2.2 The Connecting Principle . . . 179
6.3 Levels of Analysis . . . 183
6.3.1 Types of Variables and of Fallacies . . . 184
6.3.2 Levels of Analysis vs. Levels of Causation . . . 186
6.3.3 Levels of Analysis . . . 187
6.3.4 Levels of Analysis and Variation in Multilevel Models . . . 189

7 Supporting the Rationale of Variation . . . 193
7.1 Variation in Mechanist Approaches . . . 194
7.2 Variation in Counterfactuals . . . 197
7.3 Variation in Agency Theories . . . 200
7.4 Variation in Manipulability Theories . . . 203
7.5 Variation in Epistemic Causality . . . 205
7.6 Variation in Single Instances . . . 207

Concluding Remarks . . . 211
1 Objectives, Methodology, and Results . . . 211
2 The Methodological Import of Philosophical Results . . . 214

Bibliography . . . 217

Index . . . 231

Preface

Explanation is at the center of scientific research, and explanation almost always involves the discovery of causal relations among factors, conditions, or events. This is true in the social sciences no less than in the natural sciences. But social causes look quite a bit different from causes of natural phenomena. They result from the choices and actions of numerous individuals rather than fixed natural laws, and the causal pathways that link antecedents to consequents are less exact than those linking gas leaks to explosions. Here as elsewhere, the foundational issues are different in the social sciences; so a central challenge for the philosophy of social science is to give a good, compelling account of causal reasoning about social phenomena that does justice to the research problems faced by social scientists. Federica Russo has done so in this excellent volume. The book focuses on probabilistic causation and causal modelling, and Russo offers a rigorous and accessible treatment of the full range of current debates. Her central goal is to shed more light on the methods of causal modelling, and she succeeds admirably in this ambition. *Causality and causal modelling in the social sciences: measuring variations* makes an important and original contribution.

In addition to its other strengths, *Causality and causal modelling in the social sciences: measuring variations* serves as a very useful and knowledgeable survey of the current state of the field of statistical modelling of causation. Russo does a very good job of positioning her understanding of the meaning of causal modelling and causal judgments in the social sciences within the broader literature, from philosophy to statistics to sociology. The book is immersed in an expert understanding of the most recent philosophical and methodological work on these topics. The issues here are complicated and technical, and readers who are not specialists in the field will appreciate the clarity of Russo's exposition. She offers a very useful discussion of the meaning of causal statements and causal reasoning in the social sciences. She also addresses the position of "causal realism" and the view that good explanations depend on discovering or hypothesising causal mechanisms underlying the phenomena.

A special strength of the book is the approach Russo takes to problems in the methodology and philosophy of social science—what she calls the "bottom-up"

approach. She proceeds on the basis of careful analysis of specific examples of reasoning about causation in the social and behavioural sciences, and works upward to more general analytical findings about causal reasoning as it actually works in several areas of the social sciences. Her case studies are well selected and well done. And they allow the reader to see the point of the more abstract issues very concretely. This is a particularly fruitful approach to the philosophy of social science: examine in detail the kinds of concepts and reasoning that are actually put forward by working social scientists as they grapple with complex social phenomena; and then attempt to make clear philosophical sense of these intellectual practices. We should not begin with the assumption that the difficult issues of social-scientific reasoning can be resolved on the basis of *a priori* reasoning; rather, we can best contribute to solutions to these issues by engaging carefully with skilled practitioners. And this is what Russo has done in this volume.

One of the central theoretical contributions that Russo makes here is to shift the focus of discussion about causation from regularities to variation. She focuses on "causation as variation": to identify X as a cause of Y is to demonstrate that variations in X produce variations in Y. She mentions several species of variation: across time, across individuals, across characteristics, across groups, and across intervention versus observation. Russo's focus on variation is very convincing; it provides an appropriate and insightful alternative approach to framing problems for social science research and explanation. And it links findings of causal relationships in the social sciences fairly directly to the interest we have in intervening in certain kinds of social outcomes—the incidence of school delinquency, for example, or the prevalence of racial discrimination in the workplace.

This book is a significant contribution to the philosophy and methodology of social science and a worthy contribution to the *Methodos* series. Social scientists and philosophers such as Wesley Salmon, Brian Skyrms, Nancy Cartwright, James Woodward, Judea Pearl, and Stanley Lieberson have developed a very deep set of controversies and debates about the proper interpretation of causal inference. Russo's book is a substantial contribution to these debates, and Russo sheds new light on some of the most difficult issues.

Daniel Little University of Michigan-Dearborn

Introduction

1 Scope of the Book and Methodology

This book deals with causal modelling in the social sciences—*social sciences* and not natural sciences. There are two main reasons motivating this choice. Firstly, it seems to me wise to focus on one particular scientific field at a time. The issue of causality has been debated in the philosophical literature as well as in the scientific literature for a long time now, and to talk about 'causality' *tout court* seems to be quite a hopeless enterprise. Many accounts try to give very general and comprehensive understanding of causality, and knock-down objections are inevitably raised. Secondly, it might be the case that causation operates differently in nature and in society. If so, different accounts of causation would need to be developed. As a matter of fact, we do not know whether this is the case. Yet, even if causation 'materially' operates in the same way in the natural and social sciences, it is still possible that causal notions are to be characterised differently in the natural and social sciences, because of differences in their objects of investigation, their different methodologies, or both.

Social science will be broadly conceived here. I shall deal with epidemiological studies on the effects of tabagism on lung cancer, with demographical models about the impact of maternal education on child survival in developing countries, with econometric studies on the possible causal links between health conditions and socio-economic status, etc. Whilst it is quite uncontroversial that economics, sociology, or demography belong to the social sciences, there is no unanimous consensus as to whether epidemiology belongs to the social sciences or to the health sciences. Perhaps, epidemiology is located somewhere in between. There were two reasons for including it in my investigation. Firstly, epidemiology is interested in how diseases are distributed within a population or across populations, and in the factors determining this distribution; thus it is a particular aspect of society that epidemiology is interested in. Secondly, epidemiology uses many of the models I shall discuss later.

F. Russo, *Causality and Casual Modelling in the Social Sciences: Measuring Variations* 1
Methodological Prospects in the Social Sciences

More specifically, this book deals with the *epistemology* and *methodology* of causal modelling. Different queries lead to different answers, therefore epistemological questions ought not to be confused with metaphysical or methodological questions.

Metaphysics studies the nature of things. Hence, the metaphysics of causality is concerned with what in fact causality is, or what the causal relata are. Answers vary from ontic accounts, where causation is the transfer of a conserved quantity in the interaction of two causal processes, to epistemic accounts, where causation is rather a mental construct. In the past, causal relata have been understood as objects, or events (although it was unclear whether type or token events), and more recently causes have been conceived to be capacities or probabilistic propensities.

Methodology addresses the problem of scientific method. Hence, the methodology of causality aims at developing successful methods for the discovery or confirmation of causal hypotheses. For instance, algorithms to discover causal relations from large sets of data have been proposed by Glymour and his associates in Pittsburgh; the development of Bayesian Networks nowadays occupies a large part of the debate on causal models, and so on.

Last, *epistemology* addresses the problem of human knowledge. Hence, the epistemology of causality investigates how we come to know about causal relations. Epistemology tackles issues such as: What epistemic access do we have to causal relations? Under what conditions are correlations causal? Does invariance, or structural stability, or the Markov condition ensure causality?

Since explanatory scientific research crucially depends on the notions of cause and effect, it is extremely important to disentangle such notions. Moreover, there is a pervasive vagueness about the questions that motivate the development of different accounts of causality, and often objections are due to this lack of clarity. Although general and comprehensive accounts of causality are not impossible *in principle*, my view is that it is methodologically wrong to start with a wide-ranging approach, for confusions might easily arise and different notions conflate. Of course, a further level of investigation may concern how the three domains—epistemology, metaphysics, and methodology—relate to each other, and how notions proper to the social sciences relate to notions proper to the natural sciences and vice-versa. So, to restrict the research domain seems to have many advantages. In particular, it allows us to make more accurate claims about causality and, using tile after tile, it leads us to a clearer reconstruction of the large mosaic of causality.

To sum up, we will be concerned here with the epistemology of causality—in particular, with the epistemology behind causal modelling in the social sciences—and with the consequences such an epistemology has on methodology of empirical research. Within this epistemological and methodological perspective, I shall tackle the following question: *What notions allow us to infer causal relations in probabilistic models in the social sciences?* In other words, *what is the rationale of causality involved in causal modelling?*

To answer this question I choose a specific methodology of philosophical investigation which I like to call *bottom-up*. This philosophical investigation into the epistemology and methodology of causal modelling is bottom-up because I first

go over some paradigmatic case studies and over the social science methodology, and then, within that discussion, I raise relevant questions. The goal is to extrapolate significant epistemological claims about causality and meaningful precepts for methodology.

2 Structure of the Book

What is the rationale of causality in causal models in the social sciences? The answer to this question will be articulated throughout seven chapters.

Chapter 1 presents five case studies: a study on the effects of prolonged cigarette smoking on lung cancer; a model about the causal impact of maternal education on child survival; an analysis of the relations between health and wealth conditions in elderly Americans; a multilevel model of migration behaviour in Norway; an attempt to develop a theory about the phenomenon of job satisfaction. These case studies are read with the philosopher's eye. This descriptive work is motivated by the following question: what kind of notions do social scientists employ in causal analysis? Do they make use of physical notions such as processes or interactions, or rather use statistical notions such as probabilistic independence, relevance, or screening-off? The simple reading of these case studies shows that social scientists opt for a *probabilistic* characterisation instead. Also, a couple of epistemological remarks on social science methodology already arise at this stage: (i) social scientists opt for statistical notions to detect causal relations; (ii) they state causal claims in causal contexts; (iii) they look for specific variations to test; and (iv) they model causal mechanisms with statistical tools.

Because social scientists characterise causal relations probabilistically, Chapter 2 examines *probabilistic* approaches. P. Suppes' and I. J. Good's probabilistic theories of causality are analysed as the philosophical attempt to give the *scheme of reasoning* to detect causal relations. Two points are emphasised: (i) the basic concept employed in probabilistic theories is statistical relevance; and (ii) probabilistic theories make a number of assumptions about causal relations—in particular, causes are assumed to precede effects in time. Probabilistic theories, however, fail to give us a meaningful rationale of causality in the social sciences mainly because the multivariate aspect of causality in social contexts is neglected. I argue, instead, that a mature probabilistic account has to focus on casual modelling, which is the object of the following chapter.

Chapter 3 opens with section 1 presenting and discussing various types of statistical models used for causal analysis: path models and causal diagrams, covariance structure models, Granger-causality, Rubin's model, multilevel analysis, and contingency tables. This overview is intended for the reader who is not familiar with statistics; consequently, a considerable effort is made to avoid any unnecessary technicalities in favour of explanation of the concepts involved. Two issues deserve particular attention: (i) causal models make a number of assumptions—their meaning and role are spelled out; and (ii) causal models use a hypothetico-deductive

methodology to confirm or disconfirm causal hypotheses (section 2). So far Chapter 1 gives us the flavour of what social scientists do; and the first two sections of Chapter 3 accurately explain the formal tools used in causal analysis. Section 3 then offers a systematised overview of the main difficulties and weaknesses of causal modelling. The *pars construens* of the book begins here.

It is now up to Chapter 4 to develop a full epistemology of causal modelling. Section 1 answers the question: *what is the rationale of causality in causal modelling?* I argue that the notion of causality underlying causal models is *variation*: causality *as* the measure of variation. This is not tantamount to giving a definition of what causality is; rather, the rationale of variation conveys the idea that, in causal models, to test causal relations means to test variations among variables of interest. I will argue that the rationale of variation emerges from the analysis of case studies, from the investigation of probabilistic theories of causality, and it is embedded in the logic of causal models presented in Chapter 3. Furthermore, I show that the forefathers of quantitative causal analysis already employed the notion of variation, and that it is also present in contemporary philosophical accounts, although never made explicit. The rationale of variation, I argue, breaks down the received view that sees in regularity and invariance the bottom-line concepts of causality. Section 2 offers a taxonomy of variations according to different criteria: (i) variations across time, (ii) variations across individuals, (iii) variations across characteristics, (iv) counterfactual and control group variations, and (v) observational vs. interventional variations. However, this rationale cannot grant, *per se*, the causal interpretation. Section 3 then provides a thorough comparison of associational models and causal models. We shall see that the latter have a sophisticated apparatus made of statistical, extra-statistical, and causal assumptions. These assumptions, together with background knowledge, participate in the justification of the causal interpretation.

Epistemological issues ought to have an impact on methodology. Chapter 5 considers, in particular, the choice of the interpretation of probability in causal modelling. Section 1 recalls the problem and distinguishes two types of probabilistic causal claims: generic and single-case. Section 2 presents the leading interpretations of probability in order to give the reader enough background to understand the choice of the Bayesian interpretation. Section 3 argues in favour of the Bayesian framework, in particular, in favour of an objective or empirically-based one—both interpretations can account for the shift of meaning of probability in the two types of probabilistic causal claims sketched in the first section. Section 4 will close the chapter showing the net advantage in adopting the objective Bayesian interpretation on the grounds that it offers a better account for the design and interpretation of tests and for policy making.

Chapter 6 considers other consequences for methodology. Notably, it addresses the problem of modelling mechanisms and of the levels of causation. Section 1 characterises causal mechanisms in the context of causal modelling, paying particular attention to the modelling of mixed mechanisms, to the explanatory import of mechanisms, and to the difference between modelling mechanisms vs. modelling decision-making processes. Section 2 explores the problem of the levels of causation; firstly, it evaluates metaphysical accounts of the levels of causation; and secondly, it argues that this problem needs to be reformulated as a problem of levels

of analysis, which is the object of the following section. Section 3 then opens by offering a taxonomy of the variables used in causal modelling and by presenting two types of fallacy—atomistic and ecological—that might occur in causal modelling. The section then discusses the extent to which multilevel analysis and the rationale of variation succeed in accounting for the reformulated problem of the levels of analysis.

Finally, I go back to where I come from: philosophy. A number of philosophical accounts of causality have been proposed in the last several decades. Chapter 7 supports the rationale of variation by showing how this rationale is involved, or at least consistent, with them. We shall deal with mechanist and counterfactual approaches, with agency and manipulability theories, with epistemic causality and with single-case causal relations.

3 Philosophical Issues at the Back of One's Mind

In the closing paragraph of *Causalità. Storia di un modello di conoscenza* (Causality. History of a pattern of knowledge), Federico Laudisa (1999: 101) states:

> Anti-causal prophecies seem to have been clamorously disproved. Causal topics animate present-day debates on several issues, as for instance the nature of knowledge or perception, not to mention discussions on the causal effectiveness of mental states, or juridical, or teleological discussions. All in all, the century just ending forbears us from celebrating the funeral of causality: the insistence of human beings on necessarily seeing an objective connection between physical, moral, or mental events in their world retains causal thinking as an integral part of our *yearning for knowledge*. (My translation[1] and emphasis.)

What I like most in Laudisa's title—and of course throughout the book—is that *causality* and *knowledge* are side by side. We are thereby invited to a fascinating journey into the history of philosophy, from Aristotle to medieval thinkers, from the modern Hume and Kant to the contemporary Suppes and Salmon—the thread that Laudisa proposes is the evolution of the notion of cause and of the principle of causality, as a means of *knowing* the world around us.

In this section, I will not go through the entire history of philosophy, nor will I attempt any historical summary. Doubtless, historians provide more accurate reports about the views of thinkers of the past than I. But it is my intention to share with my twenty-seven readers some of the philosophical tenets that struck me the most and that, slowly, led me to the conclusion that causality still is a fashionable, fascinating, and fruitful topic in philosophy and in science, exactly because it constitutes a *pattern of knowledge*.

[1] In the original: "Le profezie anticausali sembrano essere state clamorosamente smentite. I temi causali animano i dibattiti odierni su questioni come la natura della conoscenza o della percezione, senza menzionare le discussioni sull'efficacia causale degli stati mentali, nonché quella di natura giuridica o teleologica. Nel suo complesso, il secolo che si chiude si è insomma ben guardato dal celebrare i funerali della causalità: l'insistenza degli esseri umani nel voler leggere una connessione oggettiva tra gli eventi fisici, morali o mentali del loro mondo rende il pensiero causale tuttora parte integrante della nostra aspirazione alla conoscenza".

The idea that causal knowledge is an essential feature for our understanding of the world is very old—it traces back to the time of ancient Greece. In the *Metaphysics* (I 1, 981 a 24–30), Aristotle maintains that knowledge is the *knowledge of causes*:

> But yet we think that knowledge and understanding belong to art rather than to experience, and we suppose artists to be wiser than men of experience (which implies that Wisdom depends in all cases rather on knowledge); and this because the former know the causes but the latter do not. For men of experience know that the thing is so, but do not know why, while others know the 'why' and the cause.

Aristotle makes similar claims elsewhere in the *Metaphysics* (A 1, 981 b 27, 982 a 3; A 3, 983 a 24–26): wisdom, which is the highest form of knowledge, is knowledge of certain principles. Hence, the meaning of cause in Aristotelian philosophy is connected to an ideal of science conceived as rigorous philosophical knowledge, where 'cause' is understood as 'principle' (*Metaphysics* VI 1, 1025b 5); that is to say, science is the necessary knowledge of causes or principles. It follows that, for Aristotle, to single out the cause means to single out the because, viz. its τὸ διὰ τί, and his Doctrine of Causes thus becomes a doctrine of principles qua explanatory principles. An essential feature of the Aristotelian causes is thus their explanatory power; nonetheless, causes are not reduced to explanations, nor are causes mere *subjective* entities that we use to organize our experience and knowledge. Instead, causes also have an *objective* character. That objective character seems to depend on the *intrinsic* (καθ′ αὑτό) connections that *objectively* take place in the world. That objectivity makes their explanatory power reliable. Aristotle seems to anticipate one of the most controversial themes of the contemporary debate: is causality merely a *subjective* feature of the agent's mental state? Or, is causality an *objective* feature of the world? That is, is causality merely related to subjective knowledge, or is it something objective, i.e. physically 'out there' in the world?

Let us try to follow this track. The trade-off between subjectivity and objectivity, epistemology and metaphysics, comes on to the scene again in modern times with Hume—a *locus classicus* of philosophy of causality. Hume's sceptical view is well known; thus I will not dwell upon it; rather, I shall focus on a particular aspect of Humean causality, namely causality as a *mind-dependent* notion. Precisely, the *necessity* of the causal connection becomes mind-dependent, for this necessity cannot be proved empirically nor logically.

The existence of the cause can be inferred from the existence of the effect—says Hume in the *Treatise*—thanks to the constant conjunction, i.e. the persistence in time of the association between the cause and the effect. However, the constant conjunction, which is based on contiguity and regular succession, presupposes the uniformity of nature, viz. the principle according to which nature will evolve in the future by the same modalities as in the past. But *that* assumption is not a logical truth; instead, its grounding is housed in our custom or *habit* of witnessing these regular successions. To sum up, if the necessity of the principle of causality is not empirical and does not carry any logical necessity, the only way out is to ground this principle into the psychological constitution of human nature. In the ultimate analysis, Hume's scepticism is not about causality itself; rather, his scepticism concerns

our *epistemic access* to the *necessary physical* connection between events. But his scepticism about the necessity 'in nature', so to speak, has repercussions on later thinkers. In some way Hume carries on a tradition where causality is tied to *knowledge*, or to the possibility of knowledge, or to the epistemological foundations of science. In this tradition we find, for instance, Kant and Mill.

To begin with, let us put the discussion of the principle of causality into the contest of Kant's work. As is well known, in the *Critique of pure reason*, Kant distinguishes two faculties: sensibility and understanding. Sensibility is the capacity to receive representations and is concerned with how objects are *given* to us. Understanding, instead, is the power of knowing an object through the representations we have of it; understanding is thus concerned with how an object is *thought*. *Knowledge*, according to Kant, can arise only through the union of the activities of sensibility and understanding.

Kant agrees with Hume in that the necessity of causal connections cannot be found empirically: hence, if there is necessity at all, this will be housed at the level of understanding, but not at the level of sensibility. Hume interrupted Kant's 'dogmatic slumbers', and that is how Kant began his 'speculative philosophy' (Kant, *Prolegomena*: Introduction). The Transcendental Analytic is thus that part of Transcendental Logic that deals with the elements of pure knowledge yielded by understanding; it is divided into the Analytic of Concepts and the Analytic of Principles. The former isolates the pure concepts of understanding; while the latter shows the *rules* that regiment the application of pure concepts to the objects. The principle of causality is exactly one of these rules. The Analogies of Experience state that experience is possible only through the representation of a necessary connection of perception, and, in particular, the second analogy is the principle of succession in time in accordance with the law of causality.

So, for Kant, the principle of causality allows us to determine the existence of phenomena in time; and it regains its necessity lost at the empirical level; however, this necessity is not a property of the relation between observed phenomena, because necessity can only come from a pure concept of understanding. That is to say, in line with Hume's critique, this necessity does not directly come from experience nor from logic; instead, it comes from the transcendental deduction of the categories. The principle of causality is an a priori synthetic judgment that enables us to distinguish between mere subjective successions and objective ones. While subjective successions are totally arbitrary, as for instance looking at a house from top to bottom or from left to right, an objective succession is that order in the manifold of appearance according to which, *in conformity with a rule*, that which happens follows that which precedes. Therefore, the relation of cause–effect is the condition of the objective validity of our empirical judgments.

Humean scepticism—that there are no objective necessary connections among phenomena given in ordinary experience—also deeply influenced J. S. Mill. Causality, in the Millian work, is approached from a privileged perspective: Mill investigates the import of causality in science, and particularly in the basic principles of experimental science. In our endeavour to understand and explain the world, we attempt to establish generalisations that we find by means of the experimental method.

Such inductive generalisations are based on the idea of uniformity in nature: inductive inferences from a limited number of observed instances to a generalisation about possible cases are feasible because nature is governed by laws. In particular, the law of universal causation assures us that for any sort of event there are laws which, if we search diligently enough, we will be able to discover. Guided by the principle of causality, we also discover that various rules of inference are more effective than others in generating acceptable causal beliefs.

In sum, these philosophers teach us mainly two things:

(i) Causal connections are worth exploring from an *epistemological viewpoint*.
(ii) Causal investigations have to be carried out at the *empirical level*.

Causal thinking, however, was savagely attacked from many quarters in the late 19th century and early 20th century. I have in mind, for instance, Ernst Mach in physics, Bertrand Russell in philosophy, and Karl Pearson in statistics. No doubt the most celebrated is Russell. The reader will have found this passage from "On the notion of cause" quoted several times—I will nonetheless cite the reference, as it is customary in the specialized literature nowadays (Russell 1912–1913: 1):

> The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.

Russell argues that the philosophers' concept of causation involving, as it does, the law of universal determinism that every event has a cause, is 'otiose'. The concept of causation ought to be replaced, instead, by the concept of functional relation. Mach, in *Knowledge and error* (1905), makes a similar point: causality, understood as a way of explaining phenomena, should be replaced by the concept of relation, which is instead a mere *description* of phenomena. Finally, the 1911 edition of *The grammar of science* contained a new chapter entitled "Contingency and correlation—the insufficiency of causation", where Pearson described causality as "another fetish amidst the inscrutable arcana of even modern science", and elsewhere as a "fantasy" and a "conceptual bondage". His correlation coefficient r was, from the start, intended as a 'measure of the intensity of association', and thus in his eyes as a replacement for talk of causation.

But anti-causal prophecies have been disproved—causality is neither a relic of a bygone age nor another fetish, as is proven by the fact that the issue of causality has not in any way disappeared from scientific and philosophical investigations. The *social sciences*, for instance, tackle the discovery and/or validation of causal relationships among phenomena of interest in order to understand, predict, and intervene in society. In these three moments of the scientific demarche, *knowledge of causes* seems to be, in many cases, a necessary ingredient in order to provide an explanation of social phenomena, as well as of individual behaviours. However, causal modelling needs more solid foundations. Causal models, in order to deliver reliable results, need to be accompanied by a thorough understanding of the notions involved and of the import of tests and of assumptions. It is indeed my intent to investigate causal modelling as a *pattern of knowledge*, as a means of acquiring knowledge about the social world.

4 Philosophy at the Service of Social Research

I should perhaps explain my personal interest in causality in the social sciences. With little knowledge of physics and little experience in playing billiards, I could quite easily figure out what 'physical' causality was about, although I recognized the complexity and controversial character of several issues. I did not intend to discredit causal issues in natural sciences, but—as it happened—I slowly became extremely curious in understanding what causality was, if anything at all, in the social sciences.

Different social sciences study society from different angles and perspectives. Sociology studies the structure and development of human society; demography studies variations in populations due to mortality, fertility, and migration behaviour; economics studies the management of goods and services, etc. These disciplines are very often concerned with *causal* claims. For instance, when economists examine the causes of unemployment or inflation, a possible (causal) explanation of these phenomena is that wage rates and alternative levels of government expenditures determine variations in unemployment or inflation. Some sociologists try to determine whether an adolescent's vocational aspirations are caused by her level of education, or by her parents' career achievements, or by her socio-economic environment and peer expectation. Recent developments of experimental psychology apply computational models (e.g., Bayesian networks) to understand how we come to learn about causal relations in the world. In epidemiology, scientists try to assess, for instance, the causal impact of tobacco consumption or of smoking cessation or of anti-tobacco policies on several diseases. The list might go on and on. I wanted to make sense of such causal claims. And this is how I came to investigate causal modelling in the social sciences. To be honest, the landfall to causal modelling has not been that straightforward.

The first thing that I noticed was that regularity theories of causality were clearly insufficient. As an heir of the Humean conception, regularity theories state that a token event a—i.e. the event of type A occurring at time t—causes b—i.e. the event of type B occurring at a later time t'—if and only if events of type A are always followed by events of type B. Regularity theories assume that there is a class of events of type A, that the token event a belongs to A, and that there is also a class of events of type B, the token event b belongs to. Thus, if events of type A are regularly followed by events of type B, i.e. events of type A and events of type B are steadily associated, and if events of type A temporally precede events of type B, then a is a cause of b. If an event has a cause, then it has a sufficient cause. Causal relations thus defined somehow imply that the causes determine their effects. Needless to say, the regularist account was, to me, very unconvincing. Many everyday and scientific examples of causal connections go against this view. For instance, there are various carcinogenic substances, but the exposure to any of them is not a sufficient cause of cancer; nor are carcinogenic substances necessary causes, however: a heavy smoker may never develop cancer.

Thus, a very naïve intuition came to my mind: the social world might be indeterministic. And this is why we have to abandon necessary and sufficient causes in favour of probabilistic ones. Regardless of the truth of determinism, social

phenomena are usually described in a *probabilistic* manner, at least because of our partial or incomplete knowledge of them. This lack of complete knowledge and of well-established social laws reverberates on the possibility of formulating sufficiently precise predictions or explanations. Most of the studies in the social sciences assume that the world of complex, multivariate, particularistic causal interdependencies we seek to understand is ordered in a probabilistic rather than in a deterministic fashion, making it impossible to predict precisely the particular behaviour of particular units at particular times.

I then became convinced that the concept of probabilistic causality might be more plausible and promising than the deterministic concept of cause, typical of regularity theories. According to probabilistic theories, causes do not necessitate their effects, but, to put it very roughly, causes only raise the probability of their effects. A further intuition motivating the interest in probabilistic relations was that the social world might be overdetermined, creating a perennial need to separate out the set of conditions responsible for a particular outcome from the many other conditions that might plausibly have brought it about. For instance, even if epidemiological studies reveal that smoking is a positive causal factor for lung cancer, it is not possible, for all that, to predict with certainty whether or not a particular individual will develop cancer, given that she smokes. The best we can do is to assign a certain probability to this hypothesis. Conversely, if we wish to explain why an individual developed lung cancer, given that she smokes, we will ascribe a certain probabilistic effectiveness to tabagism. Yet tabagism could not be the only causal factor, for the individual might have been exposed to other carcinogenic substances, thus causing her cancer. However, I quite soon realised that, to account for causal relations in the social sciences, probabilistic relations were not enough either.

It is commonly agreed that the problem of inferring causes from correlational data proved to be challenging and almost intractable for nineteenth-century scholars armed with only the most elementary techniques of statistical analysis. However, it is widely held today that, over the last three-quarters of a century, these problems have been overcome as a result of the enormous increase in sophistication of statistical methods. Granted, the problem still holds—namely, whether the most advanced of statistical methodologies can reliably be employed to generate causal knowledge—but *models*, and, more specifically, *causal models* seemed to me the right path to investigate further.

And this is how I came to causal modelling. The Humean instruction, however, is still borne in my mind: an investigation of causal relations has to be carried on at the empirical level, and, most importantly, we should be primarily interested in causal connections between causes and effects from an epistemological rather than a metaphysical perspective. With these instructions in mind I started reading the scientific and the philosophical literature, hunting for the rationale of causality, that is for the notion that regiments causal modelling.

Epistemological as well as methodological issues and case studies in social research will therefore be the object of this investigation. Although I shall always appeal to methodological and empirical arguments to reinforce my epistemological claims, the main point of this work remains highly philosophical in character: the

dominant paradigm of regularity and invariance has to be replaced with a *rationale of variation*. This philosophical investigation, as I shall argue in the central chapter of the book, has significant consequences both for philosophy and for social science methodology.

This work attempts two main contributions to philosophy. Firstly, the rationale of variation, and secondly, the bottom-up methodology of research through which I attain this result. Throughout the book, I will run an analysis of the methodology of research in quantitative social science. As anticipated in section 2, I will cast doubt on the received view which is based on a rationale of regularity and advance the view that causal modelling is governed by a *rationale of variation*. The variation rationale is not an intellectual intuition to be imposed on practising scientists, nor is it the result of a mere literature review. It arises from a *bottom-up* methodology of research in philosophy, namely a philosophical investigation that *starts* from the scientific practice, *within* the scientific practice raises methodological and epistemological issues, and *for* the scientific practice points to the path forward. Such a philosophical methodology is feasible and fruitful. Practising scientists interested in having a deep and critical understanding of their practice will hopefully find this book useful.

This work also attempts a contribution to social science methodology. In particular, I contend that there will be a methodological gain in taking seriously the epistemology of causal modelling that will be offered in Chapter 4. The benefit is at least threefold. Firstly, the analysis of the rich statistical apparatus of causal models positively participates in the justification of the causal interpretation of these models. Secondly, the specification of the different moments of the hypothetico-deductive methodology of causal modelling allows us to emphasise and to better characterise the central role of background knowledge. Thirdly, the rationale of variation carries on its own several advantages: (i) practising scientists will look for the right thing to test, that is variation; (ii) the rationale sheds light on other key notions of causal modelling, namely invariance, no-confounding, covariate sufficiency; and (iii) the rationale participates into the action-oriented goal of causal modelling because interventions have to be set up depending on what variational relations hold (rather than regular sequences of events).

5 Open Problems: Causal Realism, Objectivity, and Social Ontology

As I explained above, we are not concerned here with the metaphysics of causality. Nonetheless, from time to time, we will come across metaphysical problems, for instance in the discussion of extra-statistical assumptions of causal models. Also, in considering alternative accounts in Chapter 7, we will witness slight shifts in the meaning of objective and subjective. Sometimes, as in the mechanist approach, objective is synonymous for 'physical', elsewhere, as in the epistemic approach, objective rather means 'non arbitrary'. Whilst the former seems to presuppose causal realism, the latter does not.

Apparently, there are a couple of concepts that need further investigation: causal realism and objectivity. It is not my intention to solve any quarrels here, but rather to formulate the problem and to sketch plans for research. That is, this section is only intended to give the scent of possible new research.

Let us start from the question: Is causal realism ruled out within an epistemological perspective of causality? Because epistemology investigates *how* we come to know about causal relations, the possibility for causal realism is, at least in principle, left open. Are causal relations, therefore, objective? Well, it depends on the meaning of objective. One thing is to debate the objectivity of social enquiry and another thing is to debate the metaphysics of causality. Let me try to sketch two different paths of possible research. The first moves from the meaning of objective as 'physical', and second from objective as 'non-arbitrary'.

Objective, in the sense of 'physical', seems to presuppose causal realism. But what is causal realism in the first place? Causal realists usually require that (i) causation is objective, in the sense of 'physically' out there and not merely a feature of our thoughts or perceptions alone, and (ii) the relation between the cause and the effect is a necessary relation. It is also commonly agreed that causal realism is the view according to which the cause and the effect are linked by a causal mechanism. It is by virtue of this mechanism that we are entitled to say that C is a necessary or sufficient cause of the effect E, or that had C not occurred then E would not have occurred either, or that if C causes E then the conditional probability of E given C would be greater than the unconditional probability of E.

What would then causal realism in the social sciences be? In the social sciences, following this realist stance, causal realism would imply that there are *real* causal relations among social phenomena. That is to say, social causal mechanisms would be real and might be investigated by means of the normal empirical procedures of the social sciences. *En passant*, does being a causal realist necessarily imply being a social causal realist? If not, are there different ontologies for the natural and the social world?

A realist stance in social sciences would then entail that empirical methods can lead to hypotheses and theories that are (approximately) true of the social world and that those empirical procedures give us reasons to accept those hypotheses and theories. The possibility of attaining knowledge seems to rest, in other words, on the assumption of the existence of external, i.e. mind-independent, social causal mechanisms. Causal mechanisms, however, raise at least two questions. One relates to social ontology and another to explanation.

In fact, if these causal mechanisms are considered to be real, one might wonder exactly what the relata in the relations involved in causal mechanisms are. This question quite naturally leads us to questions about social ontology, because we have to understand the ontological status of the entities involved in those causal relations. Also, if there are such causal mechanisms, then they ought to participate in the explanatory endeavour of the social sciences. But what do social sciences exactly *explain*? Social sciences seem to provide general explanations of average behaviours in a population, but also singular explanations of the behaviour of individuals. So what is the logic behind such explanations? How is the relation between

average and individual behaviours to be understood? Moreover, objectivity also invokes an assertion of rational credibility for the theories advanced in the social sciences. Differently put, this a problem of *epistemic* objectivity. Epistemic objectivity can be attained, we might think, notwithstanding the *ontic* objectivity of causal relations, i.e. notwithstanding the plausibility of social causal realism.

As is well known, there are, nonetheless, several sceptical concerns about objectivity—in the sense of non-arbitrariness—in the social sciences. There seem to be no 'pure' facts, but only facts couched in one conceptual system or another; there seem to be no 'pure' observations, but rather observations couched in a theory-laden vocabulary; research projects are often guided by prior assumptions about the structure of the phenomena which shape the eventual empirical findings in an arbitrary way; etc.

So what can ensure objectivity in the social sciences? The problem of objectivity has also been reformulated as a problem of *evidence*. Evidence has been considered the benchmark of objective knowledge. But evidence, just like objectivity, is neither monolithic nor unchanging. What counts as evidence may vary from discipline to discipline, and/or from research tradition to research tradition. *En passant*, given the emergence, for instance, of an *evidence based* medicine, isn't it urgent to develop a more sophisticated concept of evidence?

Besides, objectivity also seems to make reference to the *adequacy* of models with respect to the social phenomenon at stake. To what extent does the adequacy of different methods depend on facts about *social ontology*? Are different methods in causal analysis (e.g. structural models, Bayesian nets, decision analysis) applicable in different disciplines? That is to say, is there a unity of causal inference? Would different types of analyses lead to the same results? Are different techniques applicable across different disciplines? If so, why? If not, is it a matter of fact, or might we hope that further developments will?

As the title of this last section suggests, I will not give a thorough answer to those problems. However, my attitude towards causal modelling, far from being sceptical and distrustful, is constructive. I do think that causal models are a useful and powerful tool to gain knowledge about the causal mechanisms that regiment society (see later Chapter 6), and I do hope that my investigation into causal modelling will provide the social science methodology with more solid foundations. Yet, there is a pervading scepticism about philosophers nosing into science. The practising researcher might still see philosophy as resembling Minerva's owl—who always arrives at dusk. Nevertheless, my hope is that after reading the book the practising researcher will be convinced of the urgency and pertinence of philosophical investigations into social science methodology.

Chapter 1
What Do Social Scientists Do?

Abstract This chapter investigates whether Salmon's account of causality in terms of physical processes and interactions makes justice of the type of causal claims made in social contexts. Based on some paradigmatic case studies from the social sciences (e.g. demography, epidemiology, or econometrics), it is argued that social scientists (i) use statistical causality to detect causal relations, (ii) state causal claims in causal contexts, (iii) look for specific variations to test, and (iv) model causal mechanisms by means of statistical tools.

Keywords Wesley Salmon; rationale of variation; statistical causality; physical process; mechanism; background knowledge; aleatory causality.

Introduction

This first chapter intends to settle the *status quaestionis* of the book. Section 1.1 considers different types of causal claims; some of them involve physical or biological processes, whereas others refer to socio-political or demographic mechanisms. I will discuss Wesley Salmon's aim to explicate *any* causal claims in terms of physical processes and interactions, and argue that this account does not seem to be a viable one in the social sciences.

Section 1.2, section 1.3, section 1.4, section 1.5, and section 1.6 intend to give a flavour of causal analysis in the social sciences by presenting five case studies, respectively: smoking and lung cancer, mother's education and child survival, health and wealth, migration behaviour, and job satisfaction. I will spell out the background of those studies and briefly explain the methods used to establish causal conclusions.

Finally, section 1.7 states the first methodological and epistemological morals that can be drawn from the case studies presented in the preceding sections. Four points, above all, will be underlined. Social scientists (i) use statistical causality to detect causal relations, (ii) state causal claims in causal contexts, (iii) look for

specific variations to be tested, and (iv) model causal mechanisms by means of statistical tools. Also, I will give reasons for a primary interest in the methodology and epistemology of causality, rather than in its metaphysics.

1.1 Different Causal Claims?

Smoking causes lung cancer. We also say that recession causes unemployment, and that because of the petrol shock living expenses increase. It is also commonly agreed that poverty causes delinquency, that solitude may lead to suicide, and that in developing countries maternal education has a causal impact on child survival. Where does knowledge of those causal relations come from? How do we know that smoking causes lung cancer? That maternal education is causally relevant for child survival?

Nowadays, the claim that smoking is a cause of cancer would hardly be denied. Even without deep epidemiological and medical knowledge, one can figure out the 'physical' mechanism leading from smoking to cancer. That is, even in a completely unsophisticated way, one can think that there is an actual causal process going on, and that is why Harry developed lung cancer: because he smoked. Moreover, many people have relatives who have contracted cancer, and so the generic claim 'smoking causes lung cancer' makes perfect sense. The claim makes perfect sense applied both to the individual and to the population.

Indeed, medical evidence has been accumulating for over 200 years—we now know what the link between cigarette smoking and lung cancer is. First of all, smoke inhalation damages the normal cleansing processes by which the lung protects itself from various injuries. Inhaled air is conducted to the lungs via the bronchi, which have a layer of cells on which lies a protective mucus. Hair-like cilia on these cells beat continuously and rhythmically to move the mucus constantly upwards from the lungs, and to remove any inhaled particles that has been trapped in the mucus. Smoke inhalation damages this cleansing mechanism very quickly: the cilia disappear and the only protection left is the mucus that, in order to replace the lack of cilia, thickens. As a consequence of those alterations, the lungs can no longer keep themselves clean. Thus, cancer-producing agents in the cigarette smoke, that would be removed by the movement of the cilia, now remain trapped in the mucus long enough to pass into the cells before these substances can be removed by the only cleansing mechanism left: coughing. Once inside the body, these chemicals can alter the nature of the cells slowly and progressively until cancer develops. Cigarette smoking causes cancer because, by destroying the cleansing mechanisms of the lungs, it facilitates the ill-omened action of chemicals contained in cigarettes. Other inhaled carcinogens will represent an additional hazard for the smoker once the effective cleansing function no longer protects the lungs. If we asked a physician what happened to Harry's lungs, this is possibly what she would say. Being such an effective cause, the same mechanism might operate in several individuals, and this is why one would not find the generalisation that controversial. So this causal claim

seems very plausible because there is a quite well understood causal mechanism behind it.

Nonetheless, causes within this causal mechanism may fail to be necessary and/or sufficient for their effects. Some heavy smokers may never develop cancer, some non-smokers may develop cancer, or cancer may be caused by other carcinogenic agents, for instance asbestos. Consequently, the claim ought to be slightly adjusted: in this causal mechanism, smoking makes lung cancer *more likely* to develop. Or, it is very *likely* that Harry developed cancer because of his heavy smoking. It seems that the *deterministic* conception of necessary and sufficient causes has had its day. In other words, causal mechanisms, even when they involve physical, biological or chemical processes, are better understood in *probabilistic* terms.

A staunch supporter of causal mechanisms is doubtless Wesley Salmon. If determinism is not true— i.e. if (at least some) events are not completely causally determined—then, if there is to be any sort of causality, it will have to be probabilistic. Thus, in his 1990 article, Salmon suggests that we ought to conceive of *probabilistic* causality as a generic term denoting *indeterministic* causality. Indeterministic causality, according to Salmon, can be tackled in two different ways: there is a distinction between aleatory causality and statistical causality.

Briefly put, *aleatory causality* places emphasis upon the *mechanisms* of causality, primarily uses concepts of process and interaction, and appeals to laws of nature (e.g. conservation of energy or momentum). Causal processes are the key because they furnish the *link* between the causes and the effects; causal processes intersect with one another in interactive forks, and in this interaction both processes are modified, and changes persist in those processes after the point of intersection. It is worth noting that causal processes and interactions are *physical* structures, and their properties cannot be characterised in terms of relationships among probability values *alone*.

On the other hand, *statistical causality* puts emphasis upon constant conjunction and statistical regularities and uses, above all, concepts such as statistical relevance, comparison of conditional probabilities, or screening-off relations. Those concepts can be defined in mere statistical terms—contrary to concepts of aleatory causality—without resorting to any physical notion. Statistical regularities are 'symptoms', so to speak, of causal relations; the Reichenbachian conjunctive fork gives the probabilistic structure of the causal relation (Reichenbach 1956); and the screening-off relation alerts us to possible situations in which, given the correlation between two events A and B, a third event C may be responsible for the correlation between A and B. For instance, coffee drinking is highly correlated with coronary heart diseases, but this is the case because people who drink coffee often smoke as well; we then say that smoking screens-off coffee drinking from coronary heart disease.

It is worth noting that Salmon does not deny that causes may be statistically relevant for their effects; what Salmon says is that causation does not necessarily entail statistical relevance, and that, instead, causal *processes* give the causal connection. In other words, physical connections are more fundamental than statistical relations. This is why, according to Salmon, we have a better chance of reaching an

adequate understanding of causation if we adopt aleatory causality rather than statistical causality. For instance, in the famous experiment of Arthur Compton, an X-ray photon collides with an electron. When the photon and the electron collide, both are modified. The frequency, energy, and momentum of the photon are changed, as are the energy and momentum of the electron, and these changes persist after the collision. So the interaction between a photon and an electron is a causal process. The definitions of causal interaction and of causal process are not statistical at all. They refer, instead, to the physical characteristics of the entities involved, and to their positions in space and time. These two concepts—and not the statistical concepts—provide the foundation of our understanding of causality.

But what about poverty as a cause of delinquency? Or maternal education as a causal factor for child survival? I am not questioning the plausibility of those causal statements. Indeed they are plausible, but—I ask—is aleatory causality a viable approach to causality in the social sciences? Salmon thinks it is. However, it seems to me that it is not self-evident to 'see' causal processes and interactions in social science scenarios. Consider the following example (Salmon 1990). A city needs a new bridge, and there will soon be an election for mayor, for which there are two candidates. The probability that the first will win is 0.4; the probability that the second will win is 0.6. If the first candidate wins, the probability that the bridge will be constructed is 0.3; if the second candidate wins, the probability is 0.7. Before the election the probability that the bridge will be built is 0.54. As it actually turns out, the first candidate becomes mayor and the bridge is built. The moral Salmon draws out of this story is that causal processes are more fundamental than statistical relevance relations. In fact, the victory of the first candidate is negatively relevant to the construction of the bridge (its probability decreases from 0.54 to 0.3), nonetheless there is a causal process—a complex socio-political process—that links the election to the construction of the bridge.

The moral I want to draw, instead, is quite different. I cannot restrain myself from asking: *What* is the process from the election to the construction of the bridge? How can causal interactions be uniquely identified in space and time in this case? Surely, as Salmon admits, the construction of the bridge involves complex social and political processes. But the puzzle is: To what extent are these social and political processes describable in terms of (causal) intersections and interactions typical of aleatory causality?

Consider another familiar example in the philosophical literature, also quoted in Salmon (1990). There is a very small probability that a given American male, randomly selected, will contract paralysis. There is a somewhat higher probability that an American male who has had sexual relations with a prostitute will contract paralysis, i.e. the sexual relation with the prostitute is statistically relevant to paralysis. Moreover, the probability of contracting paralysis is still higher in cases where the American male has already contracted syphilis, but in that case the sexual relation with the prostitute is no longer relevant. Processes of human physiology are involved here; suppose an American male is selected, say Harry; then the causal process from his contracting syphilis to his developing paralysis may be describable in terms of intersections and interactions of (physiological) causal processes. However, what

would the causal process be if we were to study the relation between syphilis and paralysis in the adult American *population*?

It seems that even though it might be the case that the causal process be identifiable in terms of (physical, physiological or biological) intersections and interactions, this happens at the individual-level but not at the population-level. Consider again studies on the impact of tabagism on lung cancer, or on the impact of maternal education on child mortality. One could identify the causal process from Harry's smoking to his developing lung cancer through knowledge of the effects of tabagism and knowledge of Harry's health conditions; one may even understand the causal path from Jane's level of education to her child's survival, through knowledge of typical behaviour of people similar to Jane and knowledge of socio-economic conditions Jane lives in. In this second case, however, the generalisation seems much more problematic. Obviously, generic causal claims of the latter sort cannot be true by virtue of the generalised causal claim describing the causal mechanism operating on individuals, for this very same individual mechanism looks mysterious.

Hence, the question still remains: *How* do we know? How do social scientists establish that economic distress influences marriage rates? That maternal education influences child survival? Do social scientists look for causal processes? Or do they rather adopt statistical causality? How are causal mechanisms modelled? The best way to answer those questions is to go through case studies and see how social scientists work. Thus, the first question asked in this work concerns the *methodology* of the social sciences, and the answer will be, at the very first stage, purely descriptive.

Before presenting some case studies, let me linger a bit more on Salmon's approach. In 1986 Gurol Irzik presented a paper at the Biannual Meeting of the Philosophy of Science Association (Irzik 1986) that made some quite interesting points about causal modelling, and especially he tried to bridge the gap between causal modelling and philosophical theories of causation. Although I agree with most of his claims, I still want to challenge one point.

The claim that causal relations cannot be reduced to statistical relations is uncontroversial. Irzik (1986: 19–20) agrees too, and he goes even further by claiming that Salmon's causal processes provide

> […] a suitable framework both for causal modelling and for models of causal explanation—a framework which has been anticipated and articulated by the founders of CM [causal modelling], Sewall Wright and Herbert Simon.

Irkik then quotes several places in which both Wright (1934) and Simon (1979) refer to the notion of mechanism and concludes (Irzik 1986: 21):

> Wright, a population geneticist, Simon, an econometrician, and Salmon, a philosopher, have independently showed how we can make sense of causation if we turn to processes and mechanisms rather than events. That the same notion can be meaningfully appropriated for physical, biological, and social phenomena as well as in philosophy suggests that we may have finally hit upon the correct view of causation.

I think Irzik's conclusion is due to a misleading reading of Wright and Simon. For instance, Irzik quotes Simon (1979: 79):

> Scientific inquiry is concerned not only with discovering quantitative relations between variables but also with interpreting these relations in terms of the underlying causal mechanisms that produce them.

and then says that the resemblance between Simon's and Salmon.'s view is obvious. I don't think it is. Although processes behind mechanisms are assumed, in the social sciences the mechanism itself is modelled by means of *statistical* tools, but not in terms of spatio-temporal continuous physical processes. It is worth noting that this is not a reduction of causation to statistics. This is a problem about *modelling*. I shall introduce the issue of modelling mechanisms later in this chapter and discuss it in more detail in Chapter 6.

1.2 Smoking and Lung Cancer

Epidemiological and medical studies on lung cancer do not focus only on the physical mechanisms leading from smoking to cancer, but have many other facets. The *British Medical Journal*, for instance, published a study relating trends in smoking, smoking cessation, and lung cancer in the United Kingdom (Peto et al. 2000). In this study medical evidence of the harm done by smoking is taken for granted, that is, the goal is not to ascertain whether there is a causal link at all, but rather to focus on related aspects of the same causal mechanism.

The analysis run by Peto et al., in particular, deals with national trends in smoking, in smoking cessation and in lung cancer, in comparison to the results from two large case-control studies that were conducted 40 years apart, centred on the year 1950 and later in the year 1990. The 1950 study had two main objectives: to reconfirm the importance of tobacco, and to assess the lesser effects of indoor pollution of some houses by radon. Because there has been widespread cessation of smoking, the 1990 study was able to assess the long term effects of giving up the habit at various stages. In other words, although it was already known that smoking is a cause of most deaths from lung cancer in the United Kingdom, early studies could not reliably assess the effects of prolonged cigarette smoking or of prolonged cessation. Significant results of this study concern the fact that if people who have been smoking for many years stop, they would avoid most of their subsequent risks of lung cancer.

But how did researchers get at those results? The 1950 study was conducted in London and in four other large towns; it involved interviewing as potential cases patients under 75 years of age in hospital for suspected lung cancer and, as controls, age-matched patients in hospital with other various diseases. The 1990 study was conducted in a part of southwest England. For each case a population control was obtained, selected randomly either from lists of local families or from electoral rolls. By interviews, data about smoking habits and other relevant characteristics was gathered.

How was this data then processed? How did they come up with such results from the rough data? The idea is to organize data according to different categories—e.g.

smoker, non-smoker, former smoker—across the two studies, and see how relative risks differ from one category to another and across time. Relative risks are measures of the comparative risk of developing a disease. Intuitively, the relative risk is the chance that a person exposed to certain conditions will develop the disease, compared to the chance that a person who is not exposed to the same conditions will. (For a critical note on the interpretation of risks, see Russo and Williamson (2007c: Appendix).) So, relative risks for the 1990 and 1950 studies were calculated for both men and women, and across different categories—e.g. smoker, lifelong non-smoker, etc.; those relative risks were then combined with national lung cancer mortality rates in the period 1950–1990. This serves to estimate the absolute hazards in various categories: smoker, former smoker, and non-smoker.

To put it differently, the strategy is the following: data from a 1950 study and data from a 1990 study is available; in both cases there is much information about trends in smoking, smoking cessation and the diseases people had. Thanks to that information, relative risks can be calculated for each subclass (smoker, former smoker, and non-smoker). A further and key step is to see how these relative risks differ. Calculation of relative risks for different categories allows us to assess the effects of current smoking and the effects of smoking cessation in the 1990 study. That is, the comparison of trends in smoking in the 1950 and 1990 studies allows us to estimate to what extent smoking cessation reduces the risk of developing cancer.

Results are indeed highly significant. On the one hand, Peto et al. show that people who stop smoking, even well into middle age, avoid most of their subsequent risk of lung cancer. On the other hand, they provide good evidence for a predictive claim: mortality could be reduced if current smokers give the habit up. In contrast, the extent to which young people henceforth become persistent smokers will affect mortality rates mainly in the near future.

1.3 Mother's Education and Child Survival

In 1979 *Population Studies* published what nowadays is considered to be a classical model in demography: in developing countries maternal education plays a fundamental role in child survival (Caldwell 1979).

Caldwell's study concerns developing countries: in particular, he analyses rural and urban realities in Nigeria. Former studies emphasized the role of sanitary, medical, social and political factors in influencing mortality. In analysing the impact of public health services, Caldwell[1] notes that many socio-economic factors provide very little explanation of the mortality rates, whereas mother's education turns out to be of surprising importance. Since then, evidence that maternal education plays a major role in determining child mortality has been accumulated; nevertheless, little attempt has been made to explain this phenomenon. Similar evidence has been

[1] Orubuloye I. O., & Caldwell J. C. (1975). The impact of public health services on mortality: A study of mortality differentials in a rural area of Nigeria. *Population Studies*, 29, 259–272. Quoted in Caldwell (1979: 36n).

also gathered from surveys run in Greater Bombay (1966), Ghana (1965–1966), Upper Volta (1969) and Niger (1970). In sum, the role of maternal education in infant mortality certainly deserves further attention.

Caldwell's analysis is essentially based on two surveys. The first (May–June 1973) collected data about 6,606 women aged 15–59 in the city of Ibadan (Nigeria). The second (June–July 1973) collected data about 1,499 Yoruba women aged 17 or over in Nigeria's Western and Lagos States. The plan is to analyse child mortality by levels of maternal education, in order to see whether or not the causal hypothesis—that mother's education is a significant determinant and must be examined as an important force on its own—is correct. Data is then organized into tables. One table shows, for instance, the proportion of children who died by levels of maternal age (15–19, 20–24, . . . , 45–49) and by different levels of maternal education (no schooling, primary schooling only, at least some secondary schooling). The proportion of children who died is always higher in categories of less educated women. This result is not significantly different when other factors, e.g. socio-economic status, are taken into account. This reinforces the hypothesis that maternal education has significant causal effectiveness on its own. Table 4, in fact, shows only those characteristics that might explain the relationship between child mortality—namely mother's place of residence, husband's occupation and education, type of marriage (monogamy or polygamy)—and maternal education. As Caldwell (1979: 405) puts it:

> The figures in Table 4 demonstrate that, at least in terms of child mortality, a woman's education is a good deal more important than even her most immediate environment. If any one of these environmental influences had wholly explained child mortality, and if female education had been a merely proxy for them, the CM [child mortality] index would *not have varied* with maternal education in that line of the table. This is clearly far from being the case. (My emphasis.)

Caldwell's words are quite clear: maternal education *does* have causal impact on child survival. How are the results of those tables to be interpreted? That is to say, a step forward is needed by asking *why* it is so. That is, why has maternal education a causal impact on child mortality? Education, argues Caldwell (1979: 409), mainly serves two roles: it increases skills and knowledge, as well as the ability to deal with new ideas, and it provides a vehicle for the import of a different culture. Educated women are less fatalistic about illness, and adopt many of the alternatives in child care and therapeutics that become available. Educated women are more likely to be listened to by doctors and nurses. Also, in West African households, education of women greatly changes the traditional balance of familiar relationships with significant effect on child care.

But a novel analytical framework has been proposed later on by Mosley and Chen (1984). The main difference with Caldwell's model is that Mosley and Chen urge to incorporate in the analysis both social and biological variables, thus integrating approaches from the social and medical sciences. The underlying assumption of this approach is that social and economic factors affect child survival through some biological *mechanisms* that therefore need to be spelled out in more detail. Mosley and Chen justify the choice of specific biological and socio-economic variables—such as maternal health factors or environmental factors (biological), and

traditions/norms/attitudes of individuals or household-level income/wealth (socio-economic)—by appealing to the existing literature and by a thorough examination of the role those factors play. The ultimate goal is to provide a *theoretical* framework to apply in empirical studies.

1.4 Health and Wealth

In 2003 the *Journal of Econometrics* published a study dealing with possible causal paths between socio-economic status (SES) and health (Adams et al. 2003). The links between health and SES have been the object of numerous studies; this association holds for a variety of health variables and alternative measures of SES. Notably, the general finding is that higher socio-economic status is associated with better health and longer life. In the literature, there has been considerable discussion of the causal mechanisms behind this association, but there have been relatively few natural experiments that permit causal paths to be definitively identified.

In this investigation, statistical methods are applied to test for the *absence* of direct causal links from SES to health, and from health conditions to SES. In particular, the authors examine whether changes in health in a panel are influenced by features of the historical state of SES, and, conversely, whether changes in SES in a panel are influenced by features of the historical state of health. The object of this study is a sample from the population of elderly Americans aged 70 and older. Three wave surveys have been performed between 1993 and 1998 by interviewing individuals born in 1923 or earlier and their spouses. Those surveys provide data about health (e.g. cancer, heart disease, stroke, degenerative and chronic conditions, accidents, mental conditions, etc.), socio-economic status (e.g. liquid and non liquid wealth, own residence, house condition, etc.) and background demographic information (marital status, age at interview, education, etc.).

The descriptive statistics generally show a significant association between SES and health conditions. These associations, needless to say, do not prove any causal paths nor their absence. So, the question is: How is the absence of causal paths from wealth to health and from health to wealth tested? Briefly, and very informally, two types of tests are performed. There are non-causality tests and invariance tests. Non-causality tests fit the general approach of Granger (1969); Granger-causality and its application in this case study will be presented at length in Chapter 3. To put it simply, this is the logic behind a non-causality test. The present state of a variable Y_t (causally) depends on its history (Y_{t-1}) and on the history of other variables (e.g., X_{t-1}); consequently, what we want to test is whether or not the history of X conveys information for assessing the present state of Y_t. If Y_t and X_{t-1} are independent (conditional on the history of Y), then X does not 'Granger-cause' Y. Invariance tests, instead, are used to check whether model parameters stay invariant from one panel to another.

In this study there are two hypotheses at stake: (i) no causal link from health to wealth, and (ii) no causal link from wealth to health. As a general result, the

hypothesis of no causal link from health to wealth is accepted, whereas the hypothesis of no causal link from wealth to health is rejected. The authors point out that those results apply to the population under analysis and warn us that the causal mechanisms here detected say nothing about the structure of mechanisms operating in a younger population (Adams et al. 2003: 6). In fact, they only studied elderly Americans, for whom Medicare provides relatively homogeneous and comprehensive health care at limited out-of-pocket cost to the individual. Also, because the population is retired, new health problems are not likely to affect earnings.

1.5 Farmers' Migration

In a series of papers, Daniel Courgeau. (Courgeau 2002, 2003: Introduction and Chapter 2) presents a study on farmers' migration in Norway. Migratory behaviour is one of the most important fields of study in demography but my interest in Courgeau's approach is also due to the novelty of the modelling strategy.

To study the phenomenon of migration in Norway, data from the Norwegian population registry (since 1964) and from two national censuses, respectively run in 1970 and 1980, is used. The general result is that regions with more farmers are those with higher rates of migrations. In a same region, however, migration rates are higher for non-farmers than for farmers. This can be explained by a relative shortage of non-agricultural employment which encourages those in other occupations to migrate more than farmers when looking for a new job. Multilevel models will be presented in detail in Chapter 3. Let us now have a look—very informally—at how such a result is achieved.

A simplified exposition will only consider migration behaviour with respect to a single characteristic: whether individuals are farmers or non-farmers. In a first stage we consider an aggregate model in which we seek to explain the causal effect of being a farmer on the migration rate in the population. We then regress the percentage of farmers on the percentage of migration. In this model the migration rate is a linear function of the proportion of farmers in each region. The aggregate model then shows a strong propensity of farmers to migrate. However, if we estimate a logistic regression with individual data, results then go in the opposite direction: the probability of migration for farmers is 1/3 less than other professions. In this model with individual data, the probability that an individual i in the region j is a migrant is a function of her being or not being a farmer. Then, the total expected number of migrants is the sum of probabilities of each individual in that region.

To reconciliate those opposite results we can use a different model—a multilevel model—in which regional effects play a role on the probability of migrate, i.e. we consider regional percentages of farmers having an effect on non-farmers variance. Differently put, in this multilevel model aggregate characteristics (e.g. the percentage of farmers) are used to explain an individual behaviour (e.g. migrants' behaviour).

1.6 Job Satisfaction

In the late Seventies, the *American Sociological Review* published a study on job satisfaction (Kalleberg 1977). Kalleberg analyses data from the 1972–1973 'Quality of employment survey' that was conducted by the Institute for Social Research of the University of Michigan. Data is obtained from personal interviews with 1,496 persons living in housing units within the continental United States and the District of Columbia. Questions asked in the interviews concern the perceptions the interviewees had about themselves, their job, etc.

In this paper Kalleberg conducts a comparative analysis and puts forward a theoretical framework that differs from previous explanations. In particular, Kalleberg examines how work values and job rewards act upon job satisfaction. This account of job satisfaction differs from other accounts, notably from those that (i) sought to account for the variations in job satisfaction by appealing to the personalities of workers solely, (ii) viewed job satisfaction as a function of differences in the nature of jobs people perform, and (iii) considered job satisfaction as a function of the objective properties of the job plus the motives of the individual.

Kalleberg provides clear definitions and justifications of the concepts involved, for instance 'job satisfaction' or 'job reward', and puts particular emphasis upon the ways in which values, rewards and job satisfaction interrelate. His analysis is carried out both at the theoretical level and at the empirical level. In fact, attention is devoted to spelling out the plausibility of the mechanism in sociological terms and to showing how correlations between relevant variables in the model empirically support his theoretical framework. Finally, Kalleberg discusses alternative models, but argues that if different equations were used, we could not explain as much variance as his model does, and these equations would yield empirical results that go against some of the assumptions.

1.7 Methodological and Epistemological Morals

The presentation of case studies above is meant to answer the question raised at the end of section 1.1: How do scientists come to *know* about causal relations? Up to now, the answer has been merely descriptive in character. Indeed, this was done on purpose. I went through case studies just to show what social scientists do, and particularly to show how they attempt to establish causal relations. Nonetheless, because we are not doing sociology of science, we cannot content ourselves with a mere *description* of social scientists' work. Beyond descriptive methodology we encounter two further stages. I shall call them *critical epistemology* and *normative methodology*. From now on, the reader will witness an interplay between critical epistemology and normative methodology.

Epistemological and methodological issues that we will be discussing from now on, arise both from an informed and critical analysis of the scientific practice *and* from the philosophical literature. A philosophical look at case studies, i.e. a

critical survey of the scientific practice, serves several roles. It helps in uncovering neglected aspects of the scientific practice, as for instance the role of the causal context; it helps in clarifying and disentangling causal notions used by social scientists; moreover, it helps in drawing unambiguous border lines between epistemology, methodology and metaphysics. Those morals will also lead us to sensible precepts for practising scientists throughout the book.

Epistemological and methodological issues in this closing section will turn around four main axes. Social scientists (i) opt for statistical causality in order to detect causal relations, (ii) state causal claims in causal contexts, (iii) look for specific variations to be tested, and (iv) model causal mechanisms by means of statistical tools. Let us now discuss them in this very same order.

1.7.1 Statistical Causality in the Social Sciences

By way of a reminder, the need to look directly at social scientists' work was motivated by a possible difference between causal claims that involve reasonably clear causal mechanisms and causal claims that do not. I went through five case studies, and it turned out that none of them contains concepts typical of aleatory causality in order to get an understanding of causal relations—to borrow Salmon's terminology again. Instead, statistical causality is definitively preferred. However, to prefer statistical causality does not *ipso facto* rule out mechanisms from the causal talk. Put otherwise, the question is not whether or not we aim at identifying causal mechanisms, rather, *how* do we come to identify them. Causal mechanisms are not identified through causal processes and interactions, but, according to the social scientists' practice, they are *statistically modelled*. This idea that causal mechanisms are statistically modelled will be further developed in Chapter 6.

Let us consider Caldwell's model again. What does it mean to take advantage of statistical causality to study the relation between maternal education and child survival? It is commonly agreed that causal relations participate in the explanation of a given phenomenon. Or, conversely, to explain a given phenomenon it is worth appealing to causal relations. So a way to explain the relationship between maternal education and child survival is to *model* the social system this phenomenon belongs to. In turn, modelling the social system means to consider its properties and the relationships between them. The model thus built will represent the causal system. In other words, causal modelling participates in the (re)construction of the mechanisms that generate observed phenomena. Let us now see, very informally, what the ingredients of causal modelling are.

To begin with, properties of the system are customarily represented by variables. Such variables may relate to each other in different ways. Of course, because we are interested in causality, we shall try to identify what the *causal* relations between them are. Also, because of the complexity of social systems, we should expect to deal with complex and interrelated causal connections. In fact, causal influence from one variable to another can be transmitted through *other* variables. A further

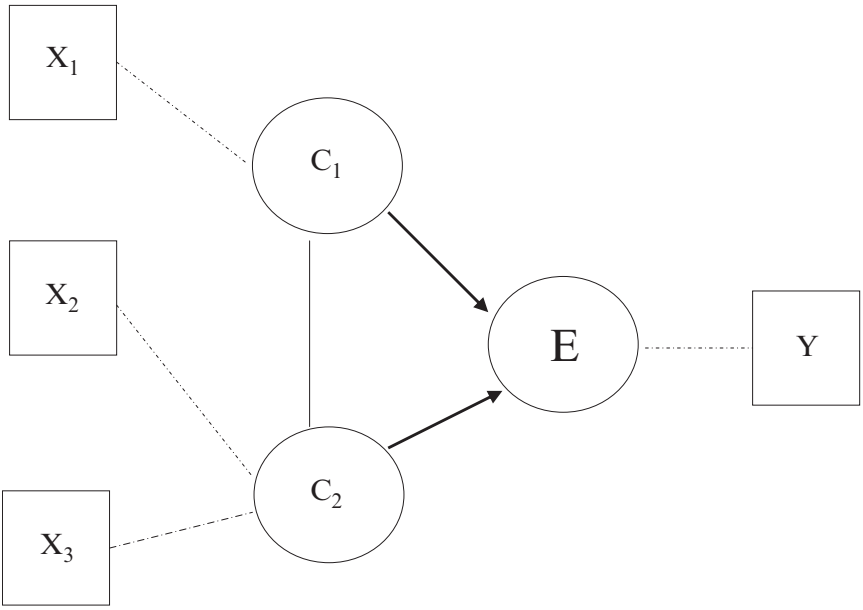

Fig. 1.1 Caldwell's model of child survival in developing countries

problem is that we do not always observe variables 'as such', rather, we have to construct, so to speak, these properties. For instance, at some point Caldwell speaks of socio-economic influence. What is socio-economic status made of? This indicator may be measured from family income, occupation of family members, area of residence, etc. Socio-economic status, operationally defined in such a way, turns out to be causally relevant for child survival. Hence, when we model properties in causal diagrams, we actually try to link indicators properly.

The causal story told by Caldwell would look more or less as in Fig. 1.1. In this causal diagram of Caldwell's model, E stands for child mortality measured by age of child at death (Y); C_1 stands for mother's education measured by years of schooling of mother (X_1); C_2 stands for socio-economic status, operationally defined by X_1, by income (X_2) and by years of schooling of father (X_3). Arrows represent causal relationships, the pointed-dashed lines denote that indicators C_1, C_2, E are measured from the variables X_1, X_2, X_3, Y; the straight line denotes that indicators C_1 and C_2 are correlated. Because of the complexity of the social system, the causal diagram is only a model, or a skeleton that displays connections of interest.

The interesting question is: *where do such (causal) connections come from*? This query might be interpreted in three different ways, namely, it might be read as a (i) metaphysical, (ii) epistemological, or as a (iii) methodological question. On the one hand, the presentation of case studies provides a very informal and *descriptive methodological* answer; a more rigorous answer will be given in section 3.1, where causal models will be presented in detail. On the other hand, the *metaphysical*

reading is merely intended here to motivate a deep *epistemological* investigation. Let me go over the argument leading from metaphysics to epistemology.

Consider again the causal system depicted in the diagram above. One can hypothesise an underlying causal mechanism made of several causal processes connecting mother's education to child survival. It would be this mechanism that produces, so to speak, the causal relations depicted in the associated graph. Whatever the *deep* causal mechanism within a social system—let us call it the black box—statistical analyses hardly enable us to get epistemic access to the *physical* processes in the black box. Nevertheless, statistical analyses *do* grant us epistemic access to properties of the system which we treat as variables. We model these properties by way of variables in the statistical model, and on this very basis causal inferences about the system under analysis are drawn. Differently put, *contra* Salmon (1990: 99), statistical causality allows us to attain *causal knowledge* of a social system.

Daniel Little (1993) draws a distinction between (i) *governing regularities*, and (ii) *phenomenal regularities*. Borrowing Little's terminology, what we would find inside the black box are the governing regularities, whereas what we attain by statistical analyses are just phenomenal regularities. According to Little, the notion of law of nature, or lawlike regularity, represents the paradigm of a governing regularity: a description of laws that generate the behaviour of a given kind of things. A lawlike regularity is a universal generalisation about empirical phenomena; it conveys necessity and may support counterfactual judgments; such regularity is grounded in the *real* physical causal properties of the entities in question. For instance, it is a governing regularity that protons and electrons are attracted by the forces described by electrodynamics. By contrast, a phenomenal regularity is a regularity of behaviour that emerges from the properties of a thing, say a social system, but does not itself determine or constrain the thing's behaviour. These regularities can be discerned through empirical investigation, and they turn out to be weak and exception-laden empirical generalisations. Social regularities are mostly phenomenal regularities and are mainly stated in probabilistic terms. One might then wonder whether our limited epistemic access to the black box is a contingent matter due to available tools. However, to answer this question goes far beyond the scope of the present work. We may leave the question open; in particular, we may leave open the possibility that in the future social sciences will also achieve knowledge of governing regularities besides phenomenal ones. However, the very idea that we aim to establish governing or phenomenal *regularities* might be challenged. Borrowing Little's terminology again, in section 4.1, I shall argue that causal models establish phenomenal (causal) *variations*, rather than regularities, and that regularity is a constraint required for the causal interpretation.

So far so good. Let us go one step further. In order to understand why the causal diagram above is a model, or skeleton, of a causal system, consider Caldwell's model again. We may summarise Caldwell's story thus:

(i) The analysis concerns populations of developing countries: in particular, data refer to the city of Ibadan (Nigeria).
(ii) Caldwell is interested in a specific feature of the social system under investigation: child mortality related to maternal education and other socio-economic factors.

(iii) In his study, Caldwell especially takes into account mother's education, because former studies focussed only on socio-economic influence and on medical or technological influence.
(iv) What does it mean to *model* these properties? Modelling the properties of a system means to study the relations between variables—causal modelling will try to establish what relations are *causal*.
(v) Caldwell's result is this: mother's education is a significant determinant in child mortality and it must be examined as an important force on its own, since maternal education cannot be employed as a proxy for general social and economic change.

Why should this story tell a *causal* story? In section 4.3 we shall examine in detail the features that guarantee the causal interpretation of those relationships. For the time being, let us focus on two reasons. One is that the whole statistical setup is built and evaluated within a *causal* context. As I mentioned above, critical epistemology helps in uncovering some neglected aspects of the scientific practice at hand. The causal context is one of them. Neither Caldwell nor the other aforementioned social scientists draw causal conclusions from a *tabula rasa*. All have, instead, a background: previous studies, evidence about the same causal relations in other countries, etc. As we shall see also in section 4.3, when such importance is bestowed on the causal context, epistemology already rules out non-sense causal connections, such as the one between the increasing number of storks and the increasing of birth rates.

The other reason is that, although *statistical* relations are used to establish *causal* links between maternal education and child survival, or between smoking cessation and cancer, we do not merely rely on the statistical correlation between variables. We want to know what the *effect* of maternal education on child survival is, that is, we want to know how, or to what extent, maternal education is responsible for changes in child mortality. In other words, we also want to *quantify* the causal impact of some variables on other variables. Of course, it might be objected that in different countries conditions are significantly different. For instance, maternal education surely does not play the same causal role in Nigeria as in the developed countries of West Europe. This is definitively true, and indeed it is the role of the causal context to fix the boundaries of the plausibility of causal claims. I shall get back to these reasons very shortly, and these issues will be further discussed in Chapter 3 and Chapter 4. Yet, in section 6.1 I will criticise Caldwell's model of the ground that the causal mechanism is *evoked* rather than modelled, and I will explain why Mosley and Chen's model take us a step forward in understanding the phenomenon.

To sum up, I will not take into account whether or not we will be able to open the black box in the future. Truly, in this investigation, I do not pretend to attain the *ontic* level and to account for the *nature* of causal relations; rather, the approach to causality will be *epistemic* here, namely causality will be discussed starting from what gives us epistemic access to causal relations. In statistical modelling, our epistemic access to causal relations necessarily passes through correlational relationships. Needless to say, correlations are not themselves sufficient to infer causation; consequently, a serious problem that causal modelling faces is what

guarantees causal inferences, which will be the object of section 4.3. In the ultimate analysis, these considerations on causal mechanisms and on the epistemic access to them quite naturally lead to investigating the epistemology and methodology of causality and causal modelling. This is how we get from metaphysics to epistemology—epistemological questions are now far more captivating.

1.7.2 Causal Claims in Causal Contexts

We thus come to discuss the second and third topics anticipated above: (ii) a causal claim is stated in a causal context, and then put forward for empirical testing; and (iii) social scientists look for specific *variations* among variables to be tested. These will occupy much of the discussion throughout the book. Extended arguments based on the current practice and methodology will be given later. In particular, the third topic will occupy the major part of Chapter 4. I will argue that the rationale of causality in quantitative social science is tantamount to testing suitable variations among variables, and this is the strongest point I want to make in this work. To provide a rationale is not to give a definition of what causality is. In this investigation, I shall look for the bottom-line concept that *guides* causal reasoning in causal modelling. At the end of section 4.3, I will also discuss why this bottom-line concept is not recognizable in the concept of regularity nor in the condition of invariance of parameters.

The second topic just mentioned will then become auxiliary, in the sense that the role of the causal context will be shown to be part of the extremely rich apparatus of causal models. In section 4.3 this apparatus will be analysed in detail. We shall see that, next to the causal context, this apparatus also has statistical, extra-statistical, and causal assumptions. I will then argue that every component has its specific role in the vindication of the causal interpretation of variations. Arguments given later in Chapter 4 will chiefly rely on a careful analysis of the social science *methodology*. Let me now give some *empirical* arguments, that is, arguments that arise from a simple reading of the scientific literature.

To say that causal hypotheses are stated in causal context means that social scientists do not start from a *tabula rasa*, so to speak. This is patently the case in the studies analysed above. The causal context in Peto et al.'s study is provided by the large evidence on smoking and lung cancer, and on the large body of literature on the same issue. Former studies already established the connection between cigarette smoking and cancer, but the new causal claim that Peto et al. wish to establish has a rather different facet: the focus is on the prolonged effect of cigarette smoking and on the effects of smoking cessation at different ages. Remember also Caldwell's study. Child mortality in developing countries has been analysed far and wide. A number of factors had already been recognized: differences in rural and urban areas had been shown to be significant, etc. In this *causal context* Caldwell suggests further investigating the role of maternal education. Adams et al. also provide extensive evidence about possible causal links between socio-economic status and health.

This causal background motivates the test for possible causal paths from health to wealth and wealth to health in the population of elderly Americans. Similarly, Kalleberg discusses previous theories of job satisfaction and on this basis justifies the development of a different causal model.

Those causal hypotheses—e.g., *maternal education influences child survival*, *smoking cessation significantly reduces the risk of cancer*—are then put forward for empirical testing. Comparisons of relative risks in Peto et al.'s study are meant exactly to accomplish that task. The same holds true for Caldwell's study: analyses of different tables test whether or not, and to what extent, maternal education has causal impact on infant mortality.

So far so good. Causal hypotheses come from background knowledge and causal claims are then stated in causal contexts. Let us now push the discussion a bit further. What is exactly this causal context? What does background knowledge precisely involve? Although intuitively clear, those concepts demand clarification. Good social science practice ought to be aware of what is within these broad notions. Background knowledge may contain information of completely different things:

(i) Similar evidence about the same putative mechanism.
(ii) General knowledge about the socio-political context.
(iii) Knowledge of physical-biological-physiological mechanisms.
(iv) Use of similar or different methodologies or data.

I shall analyse in more detail those different sources of background knowledge in section 4.3.

1.7.3 Testing Variations

The way in which I presented the case studies suggests that the logic of testing thereby employed is quite simple. We shall see, however, that things get more complex when different types of causal models are at stake. Nonetheless, and here we approach the third topic, it is worth asking *what* social scientists test.

To put it very roughly, when Peto et al. compare the risks of continued cigarette smoking with the risks of smokers who stopped at 60 years of age, at 40, etc., they actually compare levels to see whether there are *variations* or *changes* in the probability of developing lung cancer, given different smoking habits. Likewise, Caldwell analyses the *variation* of infant mortality, given different levels of maternal education. At first sight, social scientists are not primarily interested in the regularity or invariance of the relation. What they first look at is whether changes or variations take place. Let us read Caldwell (1979: 405) again:

> If any one of these environmental influences had wholly explained child mortality, and if female education had been merely a proxy for them, the CM [child mortality] index *would not have varied* with maternal education in that line of the table. This is clearly far from being the case. (My emphasis.)

A similar reasoning based on the notion of variation can be found in the case study on migrations in Norway. Witness Daniel Courgeau (2003: 75):

> In the present model it is the *variation* of these migration probabilities according to the percentage of farmers which explains the same result [i.e. why the regions with more farmers are those with higher rates of migration]. (My emphasis.)

At this point, I will be satisfied if the reader can see the intuition behind this. If not, I beg her to be patient, for the idea that causality is tested by measuring variations will be spelled out in detail in Chapter 4, and further supported throughout Chapter 7.

A couple of last thoughts. Causal hypotheses put forward for empirical testing are often called *conceptual hypotheses*. Although those claims are formulated *before* being tested, causality is not and cannot be a priori deduced from the analysis of concepts or of logical relations between the relata. Instead, causal relations are modelled—statistically modelled. The statistical model *models* the causal hypothesis, and the causal hypothesis is then put forward for empirical testing. In turn, empirical testing consists in evaluating whether the conceptual hypothesis fits the data. Results of the tests performed will then tell whether the putative causal link is confirmed or disconfirmed, i.e., statistically speaking, accepted or rejected. As we shall see in section 3.2, this corresponds to the adoption of the hypothetico-deductive methodology.

Another issue deserves attention. Even though causal models have not been deeply analysed yet, it seems quite evident that the concept of causation we are dealing with is not sharply deterministic. The regularist account (to be precise, a more sophisticated version of it), according to which causes must be necessary and sufficient for their effects, is clearly abandoned in favour of a probabilistic framework. That is, models employed in the social sciences are *probabilistic models*. It would be tempting to get into another fascinating debate, and try to solve the mystery: indeterministic causality or deterministic causality?

Unfortunately, undertaking this path would lead us well away from the main track. As we will see in section 3.1, there is a sensible use of probabilistic models for those who opt for deterministic causality and for those who advocate indeterministic causality. Nevertheless, from the perspective here endorsed, the effort will be to provide a meaningful *understanding* of the use of probabilities, rather than to solve the metaphysical quarrel once and for all. In Chapter 5, I will argue that a meaningful understanding is provided by a Bayesian framework.

1.7.4 Modelling Mechanisms

In the causal parlance of social scientists, the term *mechanism* recurs very often. Other expressions such as causal link, causal path, causal chains are usually used as interchangeably. The question that easily arises is: Is this practice correct? As I see it, those terms are not completely equivalent. In particular, ‘causal link’, ‘causal path’ or ‘causal chain’ are not completely equivalent to ‘causal mechanism’ in at least two cases. Firstly, when more than two variables are involved, and secondly,

when variables involved are of a different sort (social, demographic, economic, biological, etc.). This is patently at stake in the case study on mother's education and child survival or in the case study on health and wealth.

As for 'causal mechanism', the first thing to be clarified is that mechanism, in spite of assonance, does not rhyme with 'mechanist' but with 'structure', namely with *web of relations*. It is worth noting that even mechanistic accounts of causality—such as Salmon's account presented at the beginning of this chapter—do not necessarily commit to determinism; instead, in the social sciences causal mechanisms are, as stated above, *statistically modelled*. Chapter 6 will discuss the notion of mechanism in some detail.

At the end of this first chapter we have promising epistemological and methodological morals about the scientific practice. These morals, however, do not have philosophical depth yet. Much more work is needed in order to provide sound arguments for those insights. As we saw in the Introduction, philosophy has been debating about causality for ages, and was not behind the door with the advent of the probabilistic turn in science. So, the task of the next chapter will be to investigate whether probabilistic theories of causality advanced in philosophy can give us a meaningful rationale of causality in the social sciences.

Chapter 2
Probabilistic Approaches

Abstract This chapter examines probabilistic accounts of causality, notably P. Suppes' and I. J. Good's. The common features of probabilistic accounts are highlighted, viz. statistical relevance and temporal priority of causes, and traditional criticisms discussed. Finally, it is argued that a mature theory of causality has to start afresh by investigating causal models, as they enable us to account for the multivariate aspect of causality in the social domain.

Keywords P. Suppes; I. J. Good; probabilistic causality; twofold causality; negative causes; causal models.

Introduction

The philosophical debate on causality took a significant turn some fifty years ago, when quite well developed *probabilistic* accounts were advanced by I. J. Good and P. Suppes, in 1961–1962 and in 1970, respectively. This chapter investigates the extent to which these probabilistic approaches can be helpful in providing the *rationale* of causality in the social sciences. In a nutshell, I will argue that, although the main ideas of Good and Suppes are definitively defendable, their approaches cannot provide the rationale of causality because they are too unsophisticated.

Section 2.1 presents Good and Suppes' accounts; similarities in their approaches are underlined, and traditional criticisms discussed. In particular, I stress that both accounts employ statistical relevance as a basic notion, and assume the temporal priority of causes. In section 2.2, I focus on two categories of critiques: the first one intends to undermine probabilistic approaches from their very basis, arguing in favour of a deterministic approach; the second raises the problem of negative causes and of the levels of causation. Section 2.3 presents a first attempt to bring causal theory to maturity, namely by defining causal relations in terms of random variables and probability distributions. However, we shall see that the main objection raised against probabilistic approaches—namely that they neglect the multivariate

F. Russo, *Causality and Casual Modelling in the Social Sciences: Measuring Variations* 35
Methodological Prospects in the Social Sciences

aspect of social causality—still holds, and that the right step forward is instead to investigate causal models.

2.1 Philosophical Accounts: Good and Suppes

The reason why I discuss probabilistic approaches developed in philosophy is that before getting lost in complex formalisms, I want to draw the practising scientists' attention back to basic concepts. No wonder causality is a hard to define and elusive concept, regardless of the domain of application and of methodological preferences. For this reason good methodological accounts have to be accompanied by good conceptualizations of the notions involved therein.

In the following, we shall examine philosophical accounts in order to see whether they can help us in singling out the basic notions of causal modelling and in particular the rationale governing it. Thus, a way to look at probabilistic theories is to consider them as the philosophical attempt to give the *scheme* of the causal relation as it is analysed in scientific practice. Put otherwise, I will present probabilistic theories of causality as a *model* of scientific practice.

Probabilistic theories of causality have been developed in slightly different manners by different philosophers over the last few decades. All accounts, however, ultimately rely upon the basic intuitions of the pioneers of probabilistic theories of causality, namely I. J. Good (1961, 1962) and P. Suppes (1970). Hence, I shall focus on those two proposals. Despite considerable differences in their approaches, Good and Suppes share several common features. In particular, (i) both attempt to carry out the construction of causal relations on the basis of probabilistic relations among events, without resorting to the physical processes connecting the cause(s) to the effect(s); (ii) both employ statistical relevance as the basic concept; and (iii) both assume temporal precedence of causes. In order to grasp the main intuitions behind the probabilistic theory, I will briefly expose their accounts; I will then come back to their common points rather than their dissimilarities; and finally I will discuss the limits of these probabilistic theories of causality.

2.1.1 Good (1960–1961)

Good's approach to probabilistic causality appears to be the least familiar. One reason for this neglect may be the rather forbidding mathematical style of Good's presentation. Yet, the main features of his theory can be extracted from the heavy formalism and presented in a way that makes them easy to grasp.[1]

Good aims at providing a *quantitative* explication of probabilistic causality in terms of physical probability. Strictly speaking, however, this is not an ontological commitment to the existence of propensities, because, as Good (1983a: 197–198)

[1] Good's "Causal calculus" (Good 1961, 1962) has been reprinted in Good (1983a) and I will refer to that edition.

puts it, insofar as physical probability can be measured, it can only be done in terms of subjective probability. Although the interpretation of probability is not a matter of secondary importance for causality, at the moment it is not the main issue at stake. The interested reader may want to look at Chapter 5, where I discuss problems related to the interpretation of probability at length and defend a Bayesian, in particular an objective Bayesian, interpretation of probability.

According to Good, a quantitative explication of probabilistic causality is meant to provide a unique meaning to:

(i) The *tendency* of an event to cause another, and
(ii) The degree to which one event *actually* caused another.

Both relations (i) and (ii) refer to *events*, and although Good does not discuss the very notion of event, he at least says that they are located in space–time. The "Causal calculus" is then intended to analyse complex networks that join the initial event C (the cause) to the final event E (the effect), usually by intermediate events C_1, C_2, etc. A particular case of causal nets is the *causal chain*. A causal chain is a sequence of linearly ordered events; in Good's notation, $C = C_0$, $C_1, \ldots, C_{n-1}$, $C_n = E$.

We say that $C_0 \to C_1 \to \cdots \to C_n$ is the causal chain that connects C to E, and that the chain is composed of $n+1$ events. The main assumptions concerning causal chains are (Good 1983a: 210):

(i) Events C_i and C_{i+1} are spatiotemporally contiguous (or approximately so).
(ii) Events do not overlap too much in space and time.
(iii) All the events in the causal chain occurred, or will have occurred.
(iv) C_{i+1} started later than C_i did.
(v) C_i had a positive tendency to cause C_{i+1}.
(vi) C_{i+1} does not depend on the occurrence of any event in the chain prior to C_i, i.e. we have a Markov chain.

Some examples of causal chains are: the sequence of events leading from the event of somebody yelling "fire!" to a firing squad and the eventual death of the victim; a chemical change which affects a certain number of cells and the action of a certain number of these cells on a later biological event; the chain consisting of decay of some substance; a coin being flipped and subsequently landing on one side or another, etc.

For the purpose of this chapter, conditions (iv) and (v) deserve special attention. By condition (iv) Good explicitly *assumes* the temporal precedence of causes: causes precede effects in time by definition; however, causal priority is not enough for causation, and in fact condition (v) intervenes in requiring positive statistical relevance. Let us analyse this requirement of positive statistical relevance, viz. the *tendency of an event to cause another*, in more detail. For the sake of simplicity, we shall consider chains consisting of only two events, namely the cause C and the effect E.

The *explicatum* for the 'tendency of an event to cause another' is given by the following measure Q:

$$Q(E : C|U \cap L) \tag{2.1}$$

which is read *the tendency of C to cause E, given U and L*, where L denotes all true laws of nature—whether known or unknown—and U denotes essential physical circumstances just before the cause C started. In practice, U is not completely known, consequently, by means of the "Causal calculus", we can only make an *estimate* of the tendency of C to cause E: according to the interpretation attached to probabilities, this estimate is defined as *the subjective expectation* of the tendency to cause over all possible U's.

For brevity, some or all of what is 'given', i.e. U and L, can be omitted. As Good (1983a: 200) puts it, if we say that it is bad for eggs to be thrown in the air, we are implicitly assuming there is a law of gravitation and that there is a large gravitational body nearby, which we normally omit in everyday linguistic usage. In the same sense we can leave out the conditionalization on U and L in the explication of the measure Q, as well as in the explication of the measures S and χ which I will shortly introduce.

Good's "Causal calculus" involves a sophisticated corpus of 24 axioms and 18 theorems, which we will not explore one by one. It will suffice to pay attention to Axioms **A1–A5** that state what sort of statistical relevance Q could have, although they do not fix the precise form of Q yet:

A1. $Q(E : C)$ is a function of $P(E)$, $P(E|C)$, and $P(E|\overline{C})$;
A2. $Q(E : C)$ is a real number that may assume the value $+\infty$ or $-\infty$, or, under special circumstances, it may be indeterminate;
A3. and **A4**. $Q(E : C)$ increases continuously with $P(E|C)$, if $P(E)$ and $P(E|\overline{C})$ are held constant, and it decreases continuously as $P(E|\overline{C})$ increases if $P(E)$ and $P(E|C)$ are held constant;
A5. $Q(E : C)$ has the same sign as $P(E|C) - P(E)$.

The precise form of $Q(E : C)$ is stated much later in Good's text; according to theorem **T15**, Q is a function of $P(E|C)$ and $P(E|\overline{C})$, but it is independent of $P(E)$. In fact, Q is explicated in terms of the weight of evidence which is defined as follows:

$$W(C : E) = \log \frac{P(E|C)}{P(E|\overline{C})} \tag{2.2}$$

and it is read: *the weight of evidence concerning C provided by E*. Intuitively, the weight of evidence expresses the comparison of two possible situations: how likely is the effect E, in the *presence* of the cause C, and in the *absence* of it. The measure Q, however, is explicated by analysing the weight of evidence *against* C, if E does not happen[2]:

$$Q(E : C) = W(\overline{C} : \overline{E}) = \log \frac{P(\overline{E}|\overline{C})}{P(\overline{E}|C)} = \log \frac{1 - P(E|\overline{C})}{1 - P(E|C)}. \tag{2.3}$$

The choice of the measure of statistical relevance is somehow arbitrary, and in fact we will see later that Suppes chooses a different measure, notably, a measure that

[2] See Good (1983a: 207) for the derivation of the measure Q in terms of the weight of evidence.

depends on $P(E|C)$ and $P(E)$. The measure Q is then used to provide the measure S:

$$S(E : C|U \cap L) \tag{2.4}$$

which reads *the strength of the causal chain joining C to E, given U and L*. For a simple causal chain that consists only in the two events C and E, the strength $S(E : C)$ simply equals $Q(E : C)$, as axiom A10 states. In turn, this measure S is used to get at the measure of 'the degree to which C *actually* caused E':

$$\chi(E : C|U \cap L) \tag{2.5}$$

which reads *the degree to which C actually caused E, given U and L*. The value of χ is defined as the limit of the strength of the net joining C to E and containing all intermediate events, when the events in the causal net are made smaller and smaller (Good 1983a: 215). Good pinpoints the fact that he has not proved the existence of this limit, and that the proof, if possible, would be mathematically intricate. In the paragraph on the degrees of causation Good is indeed quite obscure, especially concerning how the measure $\chi(E : C)$ is operationally defined.

Nonetheless, this brief presentation gives us enough elements from which to draw some first conclusions about the ingredients that a probabilistic theory of causality ought to include. In particular, statistical relevance seems to be a basic concept. Besides, it is important to bear in mind that there is a substantial difference between the two measures $Q(E : C)$ and $\chi(E : C)$. To illustrate the different import of the two measures, consider the following example (Good 1983a: 216):

Holmes, Moriarty, and Watson. Sherlock Holmes is at the foot of a cliff. At the top of the cliff, directly overhead, are Dr Watson, Professor Moriarty, and a loose boulder. Watson, knowing Moriarty's intentions, realises that the best chance of saving Holmes' life is to push the boulder over the edge of the cliff, doing his best to give it enough horizontal momentum to miss Holmes. If he does not push the boulder, Moriarty will do so in such a way that it will almost certainly kill Holmes. Watson then makes the decision (event C) to push the boulder, but his skill fails him and the boulder falls on Holmes and kills him (event E).

This extravagant example is meant to show that $Q(E : C)$ and $\chi(E : C)$ cannot be identified, since C had a tendency to prevent E and *yet* caused it. We say that C was a cause of E because there was a chain of events connecting C to E, although there was no positive *tendency* of C to cause E.

Eells and Sober (1983: 38) comment *en passant* that this example expresses Good's insight that population-level causation should be distinguished from individual-level causation, and, to their knowledge, Good is the only author who had recognized the bifurcation. As we shall see later in this section, as well as in Chapter 6, the distinction of levels of causation is not just a technical detail that probabilistic theories have to incorporate. Instead, it is a genuine problem both for the epistemology and methodology of causality. The bifurcation of population-level vs. individual-level causation is concerned, on the one hand, with the identification of necessary and/or sufficient conditions that grant the corresponding causal claims;

on the other hand, the bifurcation is also concerned with the problem of how the levels relate to each other.

Bessie (1993: 15) points out that in view of the fact that Q is supposed to be a *statistical measure*, this interpretation—that the two measures Q and χ refer to two different levels, is indeed plausible. In fact, it would be correct to interpret Q as a measure of how often (i.e. in terms of relative frequencies) events of one type C bring about events of type E. If this interpretation is tenable, then statistical relevance becomes fundamental for probabilistic theories of causality. This view, however, is challenged.

In fact, Salmon (1980: 51–52) finds Good's measure Q misleading because Q is "no more nor less than a measure of statistical relevance", which indeed Good says Q is. As we saw earlier in Chapter 1, Salmon is uneasy about statistical relevance because he firmly believes that what is more fundamental in the construction of causal relations is the *physical* connection among events. Salmon's view that physical processes and interactions are the basic concepts also in the social sciences has already been challenged in Chapter 1 and therefore I will not linger on that here.

2.1.2 Suppes (1970)

In *A probabilistic theory of causality*, Suppes initially exposes his *qualitative* theory of causal relations among events. As Good (1972) correctly notices, Suppes frames his definitions in terms of measures of probability, although he does not introduce any explicit measure of statistical relevance; however, in the course of defining these causal concepts, he deploys the relations of positive statistical relevance and screening-off, but in a purely qualitative manner. In what follows, I shall recall the main definitions of Suppes' theory, and then illustrate, by means of the traditional example of the falling barometer, how these definitions work.

In the definitions and theorems of Suppes' theory, events (i) are assumed to be all subsets of a fixed probability space, (ii) are instantaneous (i.e. they are not 'chunks' of time), and (iii) their times of occurrence are included in the formal characterisation of the probability space. In the following, $P(E_{t'})$ denotes the probability of an event occurring at time t', and $P(E_{t'}|C_t)$ denotes the probability of the event E occurring at time t' given that the event C occurred at time t.

Suppes begins his discussion of causality by giving a definition of a *prima facie* cause: C is a *prima facie* cause of E if and only if C occurs before E, and C is positively, statistically relevant to E.

Definition 1. The event C_t is a *prima facie* cause of the event $E_{t'}$ if and only if

(i) $t < t'$
(ii) $P(C_t) > 0$
(iii) $P(E_{t'}|C_t) > P(E_{t'})$.

To be a *prima facie* cause is a prerequisite for being any kind of cause, other than a negative cause, which I shall define later. In order to make the next definitions

more fluid, I will omit temporal indexes. It is important to bear in mind, however, that temporal precedence of causes always belongs to the conditions stated in the definitions.

Let us now consider the definition of the *spurious cause*. The intuition behind the spurious cause is that we can find an earlier event F that accounts for the conditional probability of the effect E given the cause C. Suppes offers two definitions of the spurious cause, the second of which is stronger.

Definition 2. An event C is a spurious cause of E—*in sense one*—if and only if C is a *prima facie* cause of E, and there is an event F earlier than C such that:

(i) $P(C \cap F) > 0$
(ii) $P(E|C \cap F) = P(E|F)$
(iii) $P(E|C \cap F) \geq P(E|C)$.

Condition (ii) expresses the following idea: the occurrence of F makes C irrelevant for the occurrence of E; condition (iii), instead, says that the probability of E, conditional on C and F is greater or equal to the probability of E conditional on C, that is, if there is a causal factor responsible for the occurrence of E, this factor is F and not C. However, Suppes (1970: 23) himself is not particularly sure about condition (ii), because there seem to be arguments that favour the replacement of (ii) with the inequality $P(E|C \cap F) \leq P(E|C)$. This inequality would then express the idea that the event *really* playing a causal role for E is F, and not C, for conditioning on C makes the posterior probability of E even lower.

A different problem motivates Suppes to introduce the second definition of the spurious cause. The difficulty now concerns condition (iii) of Def. 2. Condition (iii) seems to be rather a strong constraint on the earlier event F. So, one possibility is to relax (iii) and to demand, instead, a partition[3] of the past before the spurious cause C such that, for every element in the partition, conditions (i) and (ii) of Def. 2 hold. Thus the second definition of the spurious cause is:

Definition 3. An event C is a spurious cause of E—*in sense two*—if and only if C is a *prima facie* cause of E, and there is a partition π such that for all elements F of π:

(i) $P(C \cap F) > 0$
(ii) $P(E|C \cap F) = P(E|F)$.

The move from Def. 2 to Def. 3 is significant. Def. 3 now requires that if we can observe a certain *type* of event earlier than the spurious cause C, then knowledge about this event C becomes uninformative for the prediction of the probability of the effect E (Suppes 1970: 25).

The difference between Def. 2 and Def. 3 also lies in the fact that while by Def. 2 we demand that an earlier event F *actually* exist, by Def. 3 we demand that a certain *type* of earlier event F in the partition π exist. It can be proved (Suppes 1970: 25)

[3] A partition of a sample is a collection of pairwise disjoint, nonempty sets whose union is the whole space.

that if an event C is a spurious cause in sense two, then C is a spurious cause in sense one as well. This is why Def. 3 is stronger and preferable—Def. 3 is more general in scope.

Suppes also provides definitions for the direct, indirect, supplementary, sufficient and necessary, and negative cause. Among these definitions, it will be worth mentioning the definition of the negative cause, which is defined exactly as the *prima facie* cause except that the third condition shows the inequality in the opposite direction.

Definition 4. The event C is a *prima facie* negative cause of E if and only if:

(i) $P(C) > 0$
(ii) $P(E|C) < P(E)$.

We do not need to go into further detail of the entire taxonomy of causes proposed by Suppes. The *prima facie* cause, the spurious cause, and the negative cause give us enough room for discussion, but let me first illustrate one of the most famous examples in order to show how these qualitative definitions of cause work.

The falling barometer. Let us consider the case of decreasing air pressure causing not only rain but also a falling barometer reading. The falling barometer reading is a *prima facie* cause of rain: given that the barometer is dropping, the probability that it will rain rises.

Let E denote rain, C denote a falling barometer reading, and F denote the air pressure decreasing; the probability of rain, given that the barometer reading and the air pressure are decreasing, is equal to the probability of rain given that the air pressure is decreasing, thus the condition (ii) of Def. 2 is satisfied:

$$P(E|C \cap F) = P(E|F) \tag{2.5}$$

The third condition of Def. 2 is likewise satisfied, since the probability of rain given decreasing air pressure and a falling barometer reading is at least as great as the probability of rain given a falling barometer reading:

$$P(E|C \cap F) \geq P(E|C). \tag{2.6}$$

Thus, by Def. 2, the falling barometer is a spurious cause of rain in *sense one*. However, it can be shown that the falling barometer is also a spurious cause of rain in *sense two*. In fact, let π be the partition {decreasing air pressure, non-decreasing air pressure}, then the conditions of Def. 3 are also met:

$$P(C \cap F) > 0 \tag{2.7}$$

$$P(E|C \cap F) = P(E|C), P(E|C \cap \overline{F}) = P(E|C). \tag{2.8}$$

2.1.3 Good and Suppes: Common Features

I will now focus on the common features of Good and Suppes' approaches and point out that the temporal priority of the cause and the statistical relevance requirement are at the heart of the probabilistic theories of causality.

To begin with, the *temporal priority* of the cause is assumed in both approaches. In fact, Good (1983a: 197) acknowledges his dissatisfaction with a previous work on causality where he did not make reference to time (Good 1959), hence, in the "Causal calculus", he stipulates temporal priority of causes in the definition of the causal chain. Events in causal chains are in fact spatio-temporally located, and, in particular, in the sequence of linearly ordered events in causal chains, the cause always occurs before the effect.

In Suppes' approach, causes precede their effects by definition, in other words the direction of time and thereby of the causal relation is given. As a matter of fact, Suppes acknowledges that the problem of the direction of time is still a matter for debate, yet, the source of the puzzle is that theories of classical physics permit a time reversal without discrimination. Differently put, given any model of classical theory of physics in which a concept of time occurs, the new model, obtained from the old one by reversing time, is also a model of the theory (Suppes 1970: 81). Some attempts in the philosophy of physics have been performed in order to define causality in a way that does not depend on time (Reichenbach 1956; Grünbaum 1963).

In what follows I will not linger on the problem of time in causal relations and I will assume, as is usual in much of the literature on causality—especially in the social sciences and statistics—that the analysis of interdependencies between variables as expressed in equations gives the direction of the causal relation.

Secondly, both Good and Suppes agree that *statistical relevance* is a necessary ingredient in causation. Indeed, Good incorporates the condition of positive statistical relevance into his definition of causal chains; in turn, statistical relevance serves as a measure of the strength of a causal chain. Furthermore, Suppes makes positive statistical relevance a defining condition of *prima facie* causes as well as of all other types of cause (since to be a *prima facie* cause is a necessary condition to be any other kind of cause). As Salmon (1998: 210) correctly noticed, Good and Suppes choose different measures of statistical relevance; however, for the present discussion the precise mathematical form is of secondary significance: what matters is that probabilistic approaches employ a statistical relevance measure as a basic concept.

Last, witness the following remark in Good (1983a: 221)[4]:

> This quantitative explication of causal propensity [i.e. the measure $Q(E : C)$] is basically consistent with the requirements of Suppes (1970) which, however, are only qualitative.

In his book review on Suppes' *A probabilistic theory of causality*, Good (1972) briefly recalls his "Causal calculus" and notices that Suppes has been less ambitious, for he does *not* provide a *quantitative* measure of the tendency to cause. Yet, he does not completely ignore the problem, since he recognizes that it is possible for

[4] The reference is to the 1983 collection of reprints; the paper quoted is in fact Good (1977).

a tendency to cause to be very small; in fact, he introduces a formal definition of 'ε-spurious cause' to cope with this aspect. His criterion there depends on whether a difference between two probabilities is less than a certain value ε. I am not going to dwell upon Suppes' definition of ε-causes any longer, since my prime purpose was just to stress the consistency of the two approaches.

To sum up, these few remarks allow us to extrapolate the very heart of probabilistic causality. Probabilistic theories attempt to carry out the construction of causal relations on the basis of probabilistic relations mainly among discrete events; causal relations are detected by means of statistical relevance relations among the events; and finally in all theories the temporal priority of the cause is required. I say mainly among discrete events because Suppes also gives an account of causal relations among quantitative variables (Suppes 1970: ch. 5) which I shall deal with in section 2.3, and Good (1972: 246) concedes that, although this matter was not discussed in his 1960–61 paper, he did say something about it at the 1970 Waterloo conference on the foundations of statistics. On that occasion he provided a quantitative measure for the degree to which the concomitant variables affect the dependent variables in a multinormal model.

2.2 Probabilistic Theories: Traditional Criticisms

Several sorts of criticisms have been raised against probabilistic theories (PT, hereafter). This survey is not supposed to be exhaustive, but just intends to sketch the guidelines of possible critiques. Criticisms intended to challenge Suppes' theory are, moreover, virtually also objections to Good, because both accounts share the same view on the fundamental tenets of PT.

2.2.1 Probabilistic Theories vs. Deterministic Theories

A first kind of critique intends to undermine PT at its very basis, by quarrelling over the adequacy of the *probabilistic* approach and arguing in favour of a *deterministic* one. Hesslow (1976), for instance, presents two problems for PT, and claims that those problems can be solved by the deterministic theory (henceforth, DT). According to DT, the statement '*C* caused *E*':

(i) Means that *C* is always followed by *E*.
(ii) Presupposes the presence of a sufficient condition for *E*.

The *first objection* is concerned with the requirement that causes *raise* the probability of their effects. Hesslow argues that it is possible to find causes that, instead, *lower* the probability of the effect. Consider the following example.

Contraceptive pills, pregnancy, and thrombosis. Contraceptive pills (*C*) can cause thrombosis (*T*), i.e.:

$$P(T|C) > P(T) \tag{2.9}$$

Pregnancy (P) can also cause thrombosis, i.e.:

$$P(T|P) > P(T) \tag{2.10}$$

However, contraceptive pills lower the probability of getting pregnant, i.e.:

$$P(P|C) < P(P) \tag{2.11}$$

So, contraceptive pills may lower the probability of thrombosis, because they lower the probability of pregnancy contrary to what (2.9) states, i.e.:

$$P(T|C) < P(T). \tag{2.12}$$

This example, according to Hesslow, shows a serious difficulty for the probability raising requirement of PT, whereas it does not constitute a problem for DT: no matter whether causes be positive or negative, causes are *sufficient* for their effects. In other words, this example does not trouble DT, because DT does not require causes to raise the probability of effects.

An interesting reply to Hesslow's counterexample comes from Rosen (1978), who recalls that Suppes (1970: 13) does stress that our knowledge, determination or prediction of causal relationships depends upon what information is available to us, and upon the conceptual framework relative to which we postulate causal relations. Consequently, it might be the case—as Rosen argues is the case in Hesslow's example—that the conceptual framework is inadequate and partial. Hence, relativization of probabilities to an appropriate background context would be the right way to handle the problem, and to account for the apparent inconsistency between (2.9) and (2.12) above.

Leaving aside for a moment the problem of the specification of the causal background, thorny as it may be, Hesslow's objection is ambiguous. In fact, it allows two different readings: (i) the objection is against the thesis according to which a cause will *always* increase the probability of the effect, or (ii) the objection is rather intended to pinpoint a neglect in Suppes' theory about negative causes.

In the first case, it is worth mentioning that a cause may fail to increase the probability of the effect in two ways: either by producing an improbable consequence, or by being a negative cause in spite of our expectancy. These two possibilities motivate the discussion about the levels of causation. I shall come back to this issue later in this section and also in Chapter 6. However, if Hesslow is instead arguing that Suppes failed to notice that causes may lower the probability of their effect, then he is patently wrong. We just saw that Suppes *does* provide a definition of the negative cause.

The *second objection* is that Suppes' definition of *prima facie* cause allows events that did not occur to be 'causes'. Actually, Suppes' definition does not explicitly require that both events C and E occurred; consequently, the following conditional form of a causal statement is implicitly allowed (Suppes 1970: 40): *if* the event in question occurs, *then* so and so will be the case. Hesslow's objection, in the

final analysis, is that Suppes' theory fails to distinguish causal relations between *types* of events and causal relations between *token* events. Yet, Suppes seems to be aware of the problem, and in fact he distinguishes (ibidem) between *potential* causes and *actual* causes, the latter being occurred *events* that satisfy the conditions of his theory.

This solution does not completely persuade Hesslow, though. From the fact that C may cause E (e.g. smoking causes cancer) and that C and E both occurred (e.g. Harry smoked *and* developed cancer), it does not follow that C *caused* E (e.g. Harry's smoking caused him to develop cancer). DT, instead, does not suffer this problem because—says Hesslow—the truth of the statement 'C token-caused E' only depends on the presence of those factors that together with C constitute a *sufficient* cause for E, and not on the fact that C and E actually occurred.

To plead for Suppes, although it is true that he does not explicitly distinguish between types of events and token events, when he requires that every element in partition π satisfies the conditions for the spurious cause, he is implicitly thinking—it seems to me—of token events as being of a certain *type* of events. Of course, this is my interpretation, and as such it might be wrong. But witness Suppes (1970: 79):

> A deliberate equivocation in reference between events and kinds of events runs through the earlier systematic sections of this monograph. It is intended that the formalism can be used under either interpretation [...].

So, as Rosen (1978: 610, n. 2) correctly notices, concerning the distinction between type and token events, Suppes is deliberately neutral, but he allows his definitions to stand under either interpretation. Although it is clearly a merit for Suppes' theory to be applicable at the population-level as well as at the individual-level, a few more words about the two levels and especially on how they are to be related would have been helpful.

To sum up, Hesslow's criticisms are not decisive. In the *first place*, the birth-control pills example has been widely discussed in the literature (for instance, see Rosen 1978; Cartwright 1989; Salmon 1988), and the *leitmotiv* seems to be that, as such, the depicted situation is just too simplistic: a more complete and accurate causal picture would possibly solve the difficulty raised by the inconsistency between inequalities (2.9) and (2.12). *Secondly*, if Hesslow's objection is that Suppes' theory does not account for causes that lower the probability of effects, then his objection simply does not hold. *Last*, concerning type and token events, Hesslow seems to miss Suppes' theory among quantitative properties, which is more elaborate and allows for a better characterisation of token vs. type causal relations.

2.2.2 Negative Causes and Twofold Causality

As I mentioned above, a cause may fail to increase the probability of its effect either by producing an improbable consequence, or by being a negative cause. The failure of the probability raising requirement, at least in some cases, rather than showing an overwhelming weakness of probabilistic theories, discloses the urgency

of distinguishing two levels of causation—type-level vs. token-level. Two examples are typically invoked to present difficulties related to the collapse of the probability raising requirement. Quite keenly, however, the same examples also motivate arguments in favour of twofold causality. Nonetheless, I will assert that these arguments are only partially correct, and that they are not the best strategy to support the levels of causation. Let me present the two examples first.

The sprayed plant. The notice on a can of defoliant claims that the spray is 90% effective; that is, the probability that a plant dies given that it is sprayed is 0.9, and the probability that it survives is 0.1. Suppose a plant is sprayed with the defoliant and then, improbably enough, it survives. Intuitively, the plant survived *in spite of* its being sprayed, that is, the spray did not *cause* the plant to survive.

The golf ball and the squirrel's kick. A golf ball is rolling towards the cup when a squirrel runs up and kicks it away. But, improbably enough, the ball directly falls into the cup. The squirrel's kick causes the ball to drop into the cup, even though, we may suppose, squirrels' kicks normally lower the probability of a hole-in-one.

The *first example*, due to Cartwright (1979), intends, above all, to show an asymmetry between causality and explanation. In fact, one can explain why some plants died by remarking that they were sprayed with a powerful defoliant, but this does not explain at all why some survived. The reason lies in the fact that if we are to explain an event by its cause(s), then, if the plant survives, we are inclined to say that the plant survives *despite* its being sprayed, hence this would *not* count as a (positive) cause, at least according to PT.

The *second example* is discussed in several papers (Eells and Sober 1983; Sober 1984, 1986). In this modified version,[5] it intends to show that it is plausible to claim that a cause *lowers* the probability of the effect *and* that PT is correct in its basic idea. The plant example is also supposed to provide evidence for the same statement.

Let us go one step back and try to analyse their probabilistic structure—this will help to clarify why the sprayed plant or the squirrel's kick are supposed to harm PT. For the sake of simplicity I omit temporal indexes and consider the causal chain constituted by only two events. Let E denote the event 'the golf ball drops into the cup', and suppose that the unconditional probability $P(E)$ is quite high; also, let C denote the event 'a squirrel kicks the ball'. It then seems straightforward to estimate the conditional probability of E given C lower than the unconditional probability of E; in other words, C is a *negative cause* for E, that is:

$$P(E|C) < P(E). \tag{2.13}$$

[5] For the sake of philology, the original version of this example is due to Deborah Rosen, as reported by Suppes (1970: 41); see also Rosen (1978). The story goes like this: a mediocre golfer, Jones, on his approach shot hits a tree-limb with the spectacular result that the ball is deflected directly into the cup for a birdie. In advance of play we would give Jones a low probability of making a birdie on this hole, and we would estimate the probability of making a birdie as still lower given the information that Jones hit a branch. Yet when we see the event happen, we recognize that hitting the branch was essential to the ball's going into the cup. It is worth noting that the two versions do not differ in a substantial way; in the following, I shall discuss Sober and Eells' version, simply because it has become the most popular one.

What happens next in the example is that the squirrel actually kicks the ball, which falls straight into the cup. The squirrel's kick now seems to be a *positive cause* of the birdie, that is, according to PT:

$$P(E|C) > P(E). \tag{2.14}$$

And here comes the trick. How can it be the case that the same event—namely, 'the squirrel's kick'— be a negative *and* a positive causal factor? At first sight PT is in trouble, if we want to claim that the squirrel's kick is a negative cause *and* that the squirrel's kick actually caused the ball to fall in one. Eells and Sober (1983) and Sober (1984) then argue that it is still possible to support PT, provided that we distinguish two levels of causation. In fact, this kind of situation at least suggests that type-level causal claims do not constitute a necessary condition for token-level causal claims.

Let us analyse the argument in more detail. Upper case letters C and E will denote *type*-events such as 'kicking the ball' or 'spraying the plant'; lower case letters c and e will denote *token*-events such as 'the squirrel kicked the ball' or 'the plant has been sprayed'; the arrow $\rightarrow$ will then denote the causal relation—the causal relevance can be positive or negative as indicated in the superscript of the arrow.

The contradiction between the two formulas (2.13) and (2.14) above seems to arise because even if the statement $C \overset{-}{\longrightarrow} E$ at the population-level holds—namely, 'squirrels' kicks lower the probability of birdies', it might be the case that, in a particular situation, the corresponding token causal relation simply does not hold. In fact, in the example above, the squirrel's kick actually causes the ball to fall straight into the cup, viz. $c \overset{+}{\longrightarrow} e$, although at the population-level squirrels' kicks are known to be preventatives. In other words, the same causal arrow $\rightarrow$ is made explicit in the two probability inequalities (2.13) and (2.14) which have *opposite* directions. The same may apply to the plant example.

A hasty conclusion would be that since probability inequalities at the population-level and at the token-level do not match, PT is in serious trouble. Sober and Eells' conclusion, on the contrary, is that PT is still coherent, provided that two levels of causation be distinguished. Incidentally, Sober also stresses the fact that, although both examples exhibit an apparent contradiction between population-level and token-level causal relations, in one case we are inclined to say that the squirrel caused the ball to fall into the cup, but, in the other case, we would not say that the spray caused the plant to survive. As I understand it, Sober might be suggesting that a mature probabilistic theory of causality should account for positive and negative causes as well. If my interpretation is right, Suppes' theory already fulfils Sober's hope, for Suppes *does* include negative causes in his account.

Nonetheless, Sober's remark still holds: we would say that the squirrel's kick caused the birdie, but that the plant survived in spite of its being sprayed. An interesting solution to these opposite intuitions can be found in Salmon (1980, 1984). Salmon explains the difference between the squirrel's kick and the sprayed plant by saying that the squirrel's kick started a *new process* in the trajectory of the ball, and consequently *caused* the ball to drop in. On the other hand, in case the plant

survives, spraying the plant did not start a new process, and this is why we are inclined to say that the plant survives *despite* its being sprayed.

In my view, this line of argument might be promising for naïve examples such as those we are dealing with, or even for more complicated situations encountered in natural sciences; however, as I already pointed out in Chapter 1, it is hard to see how processes that intersect and interact may fit complex scenarios such as those described in the social sciences. Instead, the statistical causality sketched in Chapter 1 seems to be a viable approach, even though the apparatus of PT is still too unsophisticated. Besides, aleatory causality can account for the difference between the squirrel's kick and the sprayed plant only at the price of a metaphysical assumption of a different causal mechanism operating at the token-level. This is how Eells (1991), for instance, regains the squirrel's kick *qua* positive cause. I shall challenge the plausibility of such metaphysical assumptions in Chapter 6.

To plead further for probabilistic theories, the squirrel's kick or the sprayed plant are not, after all, real difficulties for PT, for PT has been elaborated exactly in order to avoid the claim that causes necessitate or are sufficient to their effects. Hence, we should be glad to obtain the following results:

(i) The truth of a generic causal claim does not imply the truth of the corresponding token.
(ii) A token causal claim is not necessarily an instantiation of the corresponding generic one.

As I will spell out later in Chapter 6, I am myself very sympathetic with a twofold conception of causation, although I will offer a better argument in its favour. In particular, I shall argue that twofold causality is better understood in terms of *generic* and *single-case* rather than, respectively, population-level and individual-level or type-level and token-level. The new pair of concepts better succeeds in accounting for the scientific parlance, and, moreover, allows for the fact that individual causal claims may yet be generic.

A last remark before closing this section. PT, as it is formulated above, does not adequately mirror the scientific practice yet. This is for two reasons. The *first* is that social researchers normally do not handle events but variables, and variables convey much more information than events. An event simply occurs or does not, whereas a variable contains much information about how a given characteristic spreads across the sample, namely how a given characteristic *varies*. The *second* is that PT copes only with bivariate cases, and thus neglects the *multivariate* aspect of social causality. The first problem might be overcome by saying that although the appropriate causal language speaks the language of random variables, probability distributions, correlation coefficients, and so forth, causal relations among events are a simplified *façon de parler*. However, the second problem still remains. A first attempt to improve the probabilistic theory of causality accordingly comes from Suppes himself, as I will immediately show; we shall also see in section 3.1 how the multivariate aspect of social causality is instead taken into account in causal models.

2.3 Bringing Causal Theory to Maturity

Chapter 5 of Suppes' *A probabilistic theory of causality* is devoted to causal relations among quantitative properties. All definitions in this chapter have content similar to the corresponding definitions given earlier in Suppes' book, although they differ in the formal tool of analysis: whilst in Chapter 2 probability functions are defined over events, in Chapter 5 they are defined over random variables or quantitative properties.

Very surprisingly, the philosophical literature has not paid attention to this part of Suppes' masterpiece. As a matter of fact, there is no substantial difference between the account of causal relations among events and among quantitative properties. Nonetheless, it seems to me that the latter helps in clarifying the source of misleading arguments lingering in the literature and provides the basis for a mature probabilistic understanding of causality.

To begin with, quantitative properties are represented by random variables, namely by real-valued functions defined over a sample space, and finite families of random variables are assumed to have a joint probability distribution. Consider the definition of the *prima facie* cause. For brevity, I omit temporal indexes, but the temporal precedence of the cause is again stated in the definition.

Definition 5. The property Y is a (weak) *prima facie* cause of the property X if, and only if: for all x and y, if $P(Y \leq y) > 0$, then $P(X \leq x | Y \leq y) > P(X \leq x)$.

This condition may also be written as follows:

$$P(X \leq x \cap Y \leq y) > P(X \leq x) \cdot P(Y \leq y) \tag{2.15}$$

that states a condition of *dependence* between Y and X: the probability of a joint occurrence of Y and X is greater than the probability in the case Y and X would occur independently. This is exactly the definition of dependence as given by Lehnmann (1966).

Hence, temporal priority of the causes *and* dependence between the variables of interest seem to be the main ingredients of a probabilistic theory of causality. A step further can be taken by asking how strong the dependence between Y and X is, i.e. we wish to estimate their *measure* of association. In many ways the theoretically most important measure of association of random variables is their *covariance*. Covariance, however, is sensitive to the scale of measurement, so the *correlation coefficient* is more often used instead. Following Lehnmann again, Suppes proves the following theorem (for the proof, see Suppes 1970: 118–119), which shows that the notion of the *prima facie* cause thus defined *implies* non-negative correlation.

Theorem 1. *If Y is a prima facie cause of X, and if the variances of Y and X exist as well as their covariances, and if neither of their variances is 0, then the correlation between X and Y is non-negative.*

It is worth noting that what this theorem states is that *if* Y is a cause of X, *then* a nonnegative correlation is observed, provided that the hypotheses concerning variances and covariances are satisfied. Put otherwise, causation implies correlation, but,

as is well known, the reverse does not hold: for instance the correlation between Y and X might be spurious and vanish when a third variable is considered, or the correlation coefficient might equal zero although variables are statistically dependent. Causal relations now have *three* main ingredients: temporal precedence, dependence, and covariance between variables to *quantify* the association between them.

Analogously to Def. 5 of the spurious cause, the following definition is provided for quantitative properties:

Definition 6. A property Y is a spurious cause of X if, and only if, there is a property Z such that:

Y is a prima facie cause of X, and for all x, y, z if $P(Y \leq y \cap Z \leq z) > 0$ then $P(X \leq x)|Y \leq y \cap Z \leq z) = P(X \leq x|Z \leq z)$.

Let me now illustrate how these new definitions work. For this purpose, let us reconsider the familiar example of the barometer, using new definitions.

The falling barometer. Suppose we are interested in fluctuations of a barometer and in the number of rainy days over a certain period. Let the random variable Y denote the height of the barometer and let the random variable X denote the number of rainy days. For simplicity I omit temporal subscripts; as we would expect, Y is a *prima facie* cause of X, that is, according to Def. 6:

$$P(X \leq x|Y \leq y) > P(X \leq x) \tag{2.16}$$

indeed it is very likely to find a positive correlation between Y and X:

$$\rho(Y,X) > 0 \tag{2.17}$$

which is in accord to the theorem stated above: *prima facie* causation implies non-negative correlation. Next, to test whether or not the correlation between Y and X is spurious, we introduce a third variable Z, that indexes changes in the air pressure. Statistics then teaches us that spurious correlations disappear when we control the confounding variables. Obviously, the falling barometer does not cause rain, and in fact the correlation between Y and X vanishes when we hold the pre-existing meteorological index of air pressure fixed, that is

$$\rho(Y,X|Z = z) = 0 \tag{2.18}$$

The causal relation between Y and X is thus spurious, i.e., according to Def. 7 of the spurious cause:

$$P(X \leq x|Y \leq y \cap Z \leq z) = P(X \leq |Z \leq z) \tag{2.19}$$

As the example clearly shows, definitions for causal relations among quantitative properties are definitely consistent with those sketched above for the qualitative theory.

Elsewhere Suppes (1982) presents the general outline of his probabilistic theory of causality for quantitative properties that he first offered in his (1970). He then closes his paper with these words (Suppes 1982: 250):

> I have not said anything, of course, about the statistical theory associated with the ideas developed in general outline. That is another story and far too complex even to sketch in the present article. I do want to emphasize, however, that it is important to bring any causal theory to maturity by showing how it relates to detailed statistical theory and practice.

This is what the rest of the book is devoted to. This extension of qualitative concepts for causal relations among events to random variables seems to be a promising path. Moreover, Suppes' theory of causal relations among quantitative properties seems to better mirror the scientific practice. These new definitions have smoothly introduced the reader unfamiliar with quantitative causal analysis to more sophisticated tools, and the reader unfamiliar with philosophy to the fundamental notions of probabilistic approaches. However, statistical theory and practice take advantage of far more sophisticated tools: causal models, which will be presented in section 3.1.

Although probabilistic theories of causality have singled out two basic notions for causal analysis—statistical relevance and causal priority—they did not provide us with the rationale of causality we are hunting for. In Chapter 4, I will argue that this rationale relies on the notion of variation, and I will explain why statistical relevance and causal priority also ultimately rely on it.

A last thought before discussing causal models and the rationale of variation. In the *probabilistic* approaches we have examined, the problem of the interpretation of probability is not overtly addressed. Good only says, *en passant*, that as far as physical probability can be measured, it can only be done in terms of subjective probability, and Suppes does not mention the problem at all. So, here is the question: how is probability to be interpreted in causal analysis? In Chapter 5, I shall defend a Bayesian, notably an objective Bayesian, interpretation of probability. Also, I will try to draw the reader's attention to the methodological and epistemological consequences of different interpretations and in doing so argue for the urgency to take a stance in this respect.

Chapter 3
Methodology of Causal Modelling

Abstract This chapter presents causal models in detail. In particular, it introduces path models, covariance structure models, Granger-causality, Rubin's model, multilevel analysis, and contingency tables, by paying particular attention to the meaning of their assumptions and to their hypothetico-deductive methodology. An overview of the difficulties and weaknesses of causal models is also offered.

Keywords Causal models; path models; structural equation models; path analysis; Bayesian networks; covariance structure models; Granger-causality; potential outcome models; contingency tables; assumptions; hypothetico-deductive methodology; inductive methodology.

Introduction

Chapter 1 provided us with a description of social science practice by means of an analysis of case studies. That descriptive task led to a *critical epistemology*, where we asked relevant questions about the notions and the scheme of reasoning used in causal analysis. Further, Chapter 2 analysed the philosophical attempts to give a scheme of causal reasoning, mostly in the probabilistic approaches advanced by I. J. Good and P. Suppes. I underlined the reasons why these approaches are wanting, although ideas there developed disclosed a valuable potential. I suggested we should start afresh by a close investigation of causal models employed in the social sciences, which is the task undertaken in this chapter.

Section 3.1 is devoted to the presentation of causal models. I will introduce path models, covariance structure models, Granger-causality, Rubin's model, multilevel analysis, and contingency tables. In this overview I shall pay particular attention to two issues. On the one hand, I point out that causal models have a number of assumptions—their specific role and meaning will be spelled out. On the other hand, in section 3.2, I make clear that causal models use a hypothetico-deductive methodology to confirm or disconfirm causal hypotheses. Section 3.3 attempts to give a

F. Russo, *Causality and Casual Modelling in the Social Sciences: Measuring Variations* 53
Methodological Prospects in the Social Sciences

systematised overview of the difficulties that causal modelling faces. Those include: theorising vs. modelling, lack of theory, establishing causal laws, concepts and indicators, causes of effects or effects of causes, indetermination, complexity of causal structures, interpretation of tests, assessing various types of validity, and generalisability of results.

3.1 Methods and Assumptions of Causal Modelling

The analysis of dependence and variance and thereby of causality is a large topic in statistics; a great deal more is to be said about the standard methods of analysis, running from multiple correlation and the analysis of variance to regression models and covariance structure models. To begin with, I will explain what causal models are. In particular, I will present six models for causal analysis: (i) path models and causal diagrams, (ii) covariance structure models, (iii) Granger-causality, (iv) Rubin's model, (v) multilevel analysis, and (vi) contingency tables.

The specialized literature is highly technical and most of the time presupposes familiarity with several topics in statistics. I will therefore make the effort to present causal models avoiding any unnecessary technicalities in order to make the presentation as simple as possible for the reader who is not familiar with statistics. I shall pay particular attention to two issues: (i) the meaning of assumptions, and (ii) the hypothetico-deductive character of causal models. As a matter of fact, in the literature there is no unanimous consensus as to the categorization of causal models. The one I propose tries to cover as many methodologies as possible, by presenting models that are used in various social science disciplines, ranging from econometrics to demography, from epidemiology to quantitative sociology. Also, in section 3.1.1, I cluster under the label 'path models and causal diagrams' various types of analysis, including structural equation models, path analysis, and graphical models such as Bayesian networks, because of the common features they all share.

3.1.1 Path Models and Causal Diagrams

To begin with, a path model consists of a set of equations and/or a graph laying out the hypothesised causal links pictorially. Typically, equations and graphs are both employed. However, one can start by writing the equations and subsequently drawing the graph, or the other way round. To give a flavour of what a causal model looks like, I borrow an example from Pearl (2000: 27–28). Figure 3.1 depicts a canonical econometric model relating price and demand through two equations:

$$Q = \beta_1 P + \delta_1 I + \varepsilon_1 \tag{3.1}$$

$$P = \beta_2 Q + \delta_2 W + \varepsilon_2 \tag{3.2}$$

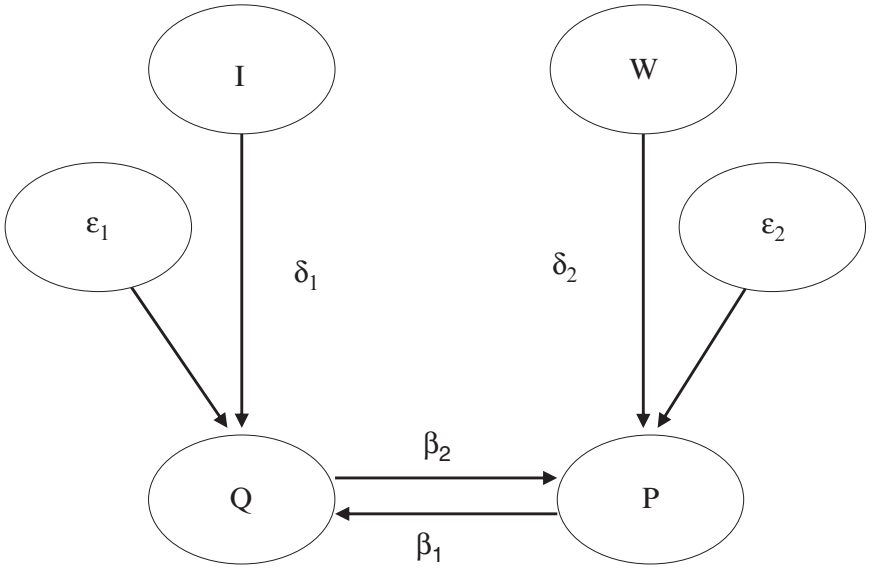

Fig. 3.1 Econometric model relating price and demand

Q is the quantity of household demand for a certain product, P is the unit price of this same product, I is the household income, W is the wage rate for producing the product, ε_1 and ε_2 are the error terms, and the Greek letters β and δ represent parameters. The first equation states that the demand depends on—or is causally determined by—the unit price of the product and the household income, while the second states that the unit price of the product depends on—or is causally determined by—the demand and the wage rate. The graph pictorially represents what equations say mathematically. Each variable is represented as a node, and relationships between variables are represented by arrows.

The Humean sceptic will immediately object that it is totally arbitrary to interpret arrows causally. Indeed, Humean scepticism with respect to causality should boost our efforts to vindicate causal inference; that is to say, Humean scepticism is a fecund starting point but a useless finishing line.

In the econometric literature, these models are customarily called 'structural equation models' (SEM). SEM have been developed so that qualitative causal information can be combined with statistical data to provide quantity assessment of cause-effect relationships among variables of interest. Let me explain further. The graph above provides *qualitative* causal information about the relation between price and demand. As arrows indicate, household income I determines quantity of demand Q, which in turn determines the price P of the product. The graph, however, does not tell us the *extent to which* I determines Q, and Q determines P. That information is instead provided by the corresponding equations: parameters β and δ *quantify* the causal impact of one variable over another.

Originators of SEM were geneticists, such as Sewall Wright, and economists, such as Trygve Haavelmo and T. C. Koopmans. In particular, Wright developed the so-called method of path coefficient or path analysis. In Wright's own words (Wright 1921: 557), path analysis is

> [...] a method of measuring the direct influence along each separate path in such a system and thus of finding the degree to which variation of a given effect is determined by each particular cause.

Let us now have a look at structural equations and associated graphs in more detail. A *graph* is made up of a set of vertices (or nodes) and of a set of edges that connect some pairs of vertices. Vertices correspond to variables in the equation, and edges denote a certain relationship that holds between pairs of variables. Each edge in the graph can be directed, i.e. marked by a single arrowhead on the edge, or undirected, i.e. unmarked. Edges can also be bidirected, to denote the existence of unobserved common causes or confounders, and in this case they are usually marked by dashed lines.

A particular type of graph is the *directed acyclic graph* (DAG). A DAG is a set of nodes and a set of arrows between some of the nodes. It is *directed* because the direction of the arrows is fixed, and *acyclic* because by following the arrows it is not possible to run into a cycle, that is, self-loops are not admitted. The corresponding structural model is usually called recursive, or Markovian.

DAGs have been used to represent causal or temporal relationships and came to be known also as *Bayesian Networks*, a coinage due to Judea Pearl (1995). Those graphs are called 'Bayesian' to emphasise (i) the subjective nature of the input information, (ii) the reliance on Bayesian conditionalization as a basis for updating information, and (iii) the distinction between causal and evidential modes of reasoning (Pearl 2000: 14). Bayesian Networks often make free use of the terminology of kinship—e.g., parents, children, descendants, ancestors—to denote various relationships in a graph.

A simple form of a *structural equation* is the following:

$$Y = \beta X + \varepsilon \tag{3.3}$$

where Y represents the putative effect, and is called the response variable; X represents the putative cause and is called the explanatory variable. β is a parameter, also called 'path-coefficient' in path analysis, and ε represents errors or disturbances or unmeasured factors.

The interpretation of the βs and εs is quite problematic. Let's start from the less difficult. Errors are susceptible of two different interpretations. According to the first one, errors represent our lack of knowledge, and consequently the structural equation represents a *deterministic* causal relation. If we had complete knowledge, we could assess the causal impact of X on Y *precisely*. Due to partial information, probabilities come into the equation exactly through this *epistemic* indeterminateness due to the errors. According to the second interpretation, errors might be interpreted as genuine chancy elements: in this case indeterminateness is rather *ontic*, that is to say, causal relations are thought to be genuinely indeterministic. Normally, philosophers

prefer the first option, whereas statisticians rather hold a stochastic representation of the world. Possibly, their claim is not a commitment to indeterminism, but just a way to say that causal relations have to be probabilistically modelled.

The causal interpretation of parameters in the equation is matter of vivid debate in the literature. As a matter of fact, originators of SEM did not have doubts about the causal interpretation of the βs—indeed, this is what path analysis was developed for, as Pearl (2000: 135–138) correctly points out. The coefficient β is meant to quantify the causal impact of X on Y. β is defined as the product of the correlation coefficient r and the ratio between the variances σ_X and σ_Y of X and Y respectively:

$$\beta = r\frac{\sigma_X}{\sigma_Y}. \tag{3.4}$$

The correlation coefficient r between X and Y is defined as the ratio between the covariance σ_{XY} of X and Y, and the product of their standard deviations:

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}. \tag{3.5}$$

It is worth noting that if the causal interpretation is adopted, the equality sign in the equation ceases to have an algebraic meaning, which would allow us to rewrite the equation (3.3) as:

$$X = \frac{(Y - \varepsilon)}{\beta}. \tag{3.6}$$

This formulation, however, has no statistical meaning for prediction nor for causality. In fact, the structural equation does not simply say that Y is equal to the algebraic sum of X and the errors ε. Instead, much information is conveyed about the *structure* of the causal relation itself in the structural equation. Because the structural equation represents a causal relation, we are allowed to predict and explain Y from X, but it is not straightforward to read the equation the other way around. This is due, among other things, to assumptions about the asymmetry of causation and causal ordering of variables. I shall come back to the validity of the causal interpretation of structural equations later in section 4.3.

More realistic frameworks employ equations with more than one explanatory variable, and *systems* of equations as well. Consider the following observational study in demography (Wonnacott and Wonnacott 1990: § 13.6). In a fertility survey, the following variables are observed for each woman: number of children (Y), age (X_1) at which women have Y children, education (X_2), and age at marriage (X_3). The relations between these four variables can be expressed in the following three equations:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_1 \tag{3.7}$$

$$X_3 = \beta_{13} X_1 + \beta_{23} X_2 + \varepsilon_2 \tag{3.8}$$

$$X_2 = \beta_{12} X_1 + \varepsilon_3 \tag{3.9}$$

and displayed as in Fig. 3.2.

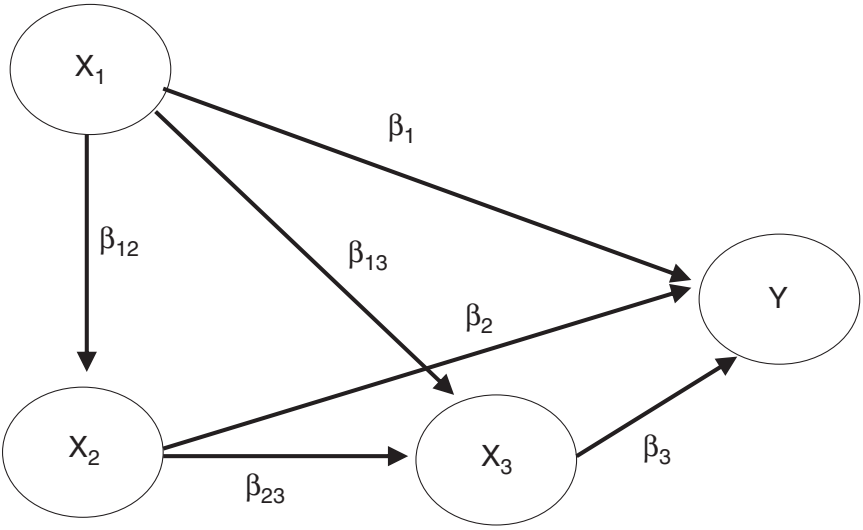

Fig. 3.2 Graph for a fertility survey

What path analysis suggests is that the total effect of one variable, say X_1, on a later variable, say Y, is defined as the change occurring in Y when X_1 changes one unit, taking into account all the changes in the intervening variables between X_1 and Y. The total effect is the sum of all paths following the arrows from X_1 to Y. In Wright's words (Wright 1921: 567):

> If a path coefficient is assigned to each component path, the combined path coefficient for all paths connecting an effect to a remote cause equals the sum of the products of the path coefficients along the paths.

Causal models also rest on a large set of assumptions. I will go through them, spelling out their meaning and their importance. Later in section 4.3 more will be said about assumptions of models, and, in particular, it will be shown that some of them play a specific role in guaranteeing the causal interpretation.

(a) Linearity and normality. The equations we came across above are linear in the parameters and in the explanatory variable X, i.e. no parameter appears as an exponent or is multiplied or divided by another parameter; the response variable Y also appears only in the first power. It is also customary to assume that errors follow the normal (Gaussian) distribution.

(b) Separability and additivity. These two assumptions claim that the response variable Y is the sum of two separated components: the explanatory variables Xs and the errors ε. This gives the 'structure' of the causal relation. Thanks to additivity and separability we are allowed to compute the total effects of the Xs on Y as the sum of the path coefficients along each path from the Xs to Y. Also, the distinction of the two separable components sheds light on the fact that, regardless of the preferred

interpretation of errors, we can specify the amount of the total effect that we are able to control, and the amount that, instead, escapes our control. The causal impact escaping our control is due to errors.

(c) Stability or invariance. Structural parameters are assumed to be stable or invariant across different panels or across interventions (whether hypothetical or performed). This means that structural parameters have to show a certain numerical stability. Two types of stability are worth distinguishing: (i) Stability of the distribution of the *population*—this concerns observations of values of a same characteristic for different individuals or units in the sample; (ii) Stability of the distribution of the *individual*—this concerns observations of various values of a same characteristic for the same individual or unit. It is worth noting that structural parameters do not necessarily represent average effects, because the parametrisation is arbitrary. Yet, when they represent average effects, stability conveys the idea 'same cause, same effect': that is, the causal factor is supposed to operate homogenously in the population of interest.

(d) No-measurement error. Variables are assumed to be accurately measured. This conveys the idea that if the expectation of errors is zero, then the causal effect Y is simply given by the sum of the explanatory variables Xs. One problem with this assumption is that many social variables are constructs measured by 'proxies', and hence they are measured only by means of other relevant variables. In practice, this assumption is too strong and often relaxed.

(e) Assumptions about errors. Errors are not correlated between themselves, nor with the explanatory variables (also called 'covariates'). If this assumption of non correlation holds, it means that all the relevant covariates have been taken into account in the model. The failure of this assumption suggests that, instead, a relevant covariate has been neglected.

(f) Homoscedasticity, or constant variance of variables. All variables are assumed to have the same variance. A situation of heteroscedasticity may be caused by the non-normality of one of the variables. In one such case the analysis loses in precision but is not definitively jeopardized. Standard techniques of transformation allow us to rectify heteroscedasticity.

Standard theorems prove that in simple linear structures if assumptions on errors are satisfied, then least square techniques for estimation give the best linear, unbiased estimates. Under the assumption of independence between X and ε, estimation is done by ordinary least square (OLS) techniques. Relaxing this independence assumption leads to generalised least square (GLS) estimators for the parameters. Also, to side step the restriction given by assumptions about the distribution of variables, distribution-free estimation methods have been developed.

A good path model—i.e. a model that describes a *causal* relation *correctly*—allows us to explain Y by means of the Xs. Notably, what we try to *explain* is the variance of Y by ascribing portions of this variation to each of the independent variables. An indicator of the goodness of fit of a model is then the *coefficient of determination* r^2, which is defined as the squared coefficient of correlation r and ranges between

0 and +1. The coefficient of determination is interpreted as the proportionate reduction of the total variation of Y by means of the explanatory X variables. Larger values of r^2 indicate that the Xs account for the total variation in Y. Under the assumption that *all* relevant explanatory variables have been taken into account, if r^2 is not the unity the residual ε quantifies the variance of Y that is left unexplained and that then escaped our control. However, it is worth noting that r^2 measures the goodness of fit, not the validity of the model itself. Therefore the explanatory power of a causal model cannot depend only on the numerical value of r^2—see also section 6.1.3 on this point.

Finally, exogeneity deserves a separate note. Exogeneity conditions have been developed mainly in the econometric literature. So far, there is no unanimous consensus as to the precise definition and interpretation of such a concept. In the early econometric literature, exogenous variables were variables 'not explained' in the model itself. For instance, Klein (1974: 13) defined exogenous variables thus:

> Exogenous variables are either a set of known numbers, that is, fixed covariates without a probability distribution, or random variables that have a marginal probability distribution with parameters that are not elements of the system being estimated.

However, exogeneity has been later on interpreted as a condition of separability of inference, when we operate a marginal-conditional decomposition over an initial joint distribution. See the locus classicus Engle et al. (1983), but also Russo et al. (2006) and Mouchart et al. (2008); for a good introduction to issues concerning exogeneity see also Hoover (2001: ch. 7).

3.1.2 Covariance Structure Models

Models for the analysis of covariance structures, also called covariance structure models (henceforth CSM) attempt to explain the relationships among a set of observed variables in terms of a generally smaller number of unobserved variables, also called latent variables. The goal of CSM is to draw conclusions about the empirical validity or non validity of causal hypotheses. In fact, the fundamental supposition underlying CSM is the *prior* formulation of the causal hypothesis. For a very good exposition of CSM, see Long (1983a, b).

Formally, CSM consist of an analysis of the covariances of the observed variables in two conceptually distinct steps:

(i) A *measurement model* linking observed variables to unobserved variables.
(ii) A *structural model* relating the unobserved variables.

The measurement component (i) of CSM consists of a pair of confirmatory factor models (CFM). These factor models explain the covariation in a set of observed variables in terms of a smaller number of common factors. The idea behind CFM is that some variables of theoretical interest cannot be directly observed. However, information about them can be obtained indirectly from their effects on observed variables. In other words, this is a procedure for uncovering a smaller number of

latent variables by studying the covariation among a set of observed variables. In turn, relations between those latent variables are analysed in the structural equation model, which is the *second* component of CSM.

Thus, CSM is the simultaneous specification of the factor model and of the structural equation model. Statistically speaking, the task is to explain the interrelationships among the observed variables as indicated by the *covariances* among them, in terms of the relationships among the unobserved variables depicted in the structural equations.

It is worth noting that what specifies *causal relations* in CSM is the structural equation model; in fact, structural equations describe the assumed causal structure of the mechanism being modelled. On the other hand, the measurement model allows us to estimate latent variables from observed variables, regardless of the structural relations among the latent variables. Hence, *these* structural relations are of greater theoretical interest. Afterwards, formal tests of adequacy between the model and the empirical data *plus* statistical signification of coefficients, permit us to conclude to the validity or non validity of the postulated causal structure at the unobservable level.

Specific assumptions are normally made in CSM:

(i) In the measurement model: unobserved variables and errors are uncorrelated.
(ii) In the structural model: explanatory variables are not correlated with errors.
(iii) Errors in the measurement model and errors in structural equations are uncorrelated. This is necessary in order to combine the measurement and structural models.
(iv) All variables are normalized.[1]

3.1.3 Granger-Causality

Granger-causality is usually employed in time-series data sets, namely in data sets where variables are observed over a certain lapse of time. More precisely, Granger-causality is concerned with the correlation of variables with histories of other variables. In other words, to evaluate causal relations between two time series, we check whether the prediction of one series could be improved by incorporating information about the other. This approach was formalized by C. W. J. Granger (1969).

Simply put, if X is a Granger-cause of Y, then one can forecast the current or future value of Y by taking the history of past values of X into account, in addition to past values of Y. Granger-causality, however, is defined through its negation, that

[1] Variables are normalized or standardized when values in different variables are adjusted in order to make them comparable. In practice, this means subtracting each value from some reference value (typically, the sample mean) and dividing it by the sample standard deviation. This transformation will bring all values, regardless of their distribution and original scale of measurement, into compatible units as if they were drawn from a distribution with a mean of 0 and a standard deviation of 1.

is, *X does not cause Y* if the conditional distribution of the effect *Y only* depends on the past history of *Y* itself and *not* on the past history of the putative cause *X*.

Granger-causality is a linear procedure, so, basically, it relies on the same assumptions as other linear methods, like those discussed above. In order to grasp the probabilistic reasoning behind Granger-causality I will introduce some new notation that, nevertheless, makes Granger's formalism easier to follow.

I shall use the capital letters $\mathcal{X}, \mathcal{Y}$ to denote the entire history of the variables *X* and *Y*, respectively:

$$\mathcal{X}_t = (X_t, X_{t-1}, X_{t-2}, \ldots) \tag{3.10}$$

$$\mathcal{Y} = (Y_t, Y_{t-1}, Y_{t-2}, \ldots) \tag{3.11}$$

I shall also use the symbol $\perp\!\!\!\perp$ for conditional independence. So, $X \perp\!\!\!\perp Y$ reads 'the variable *X* is independent of the variable *Y*', and $X \perp\!\!\!\perp Y|Z$ reads 'the variable *X* is independent of *Y*, conditional on Z'.

Consider now the histories $\mathcal{X}, \mathcal{Y}$ of the variables *X* and *Y*, respectively. Consider also the joint probability distribution of the variables *X* and *Y*, conditional on their histories:

$$P(X_t \cap Y_t | \mathcal{X}_{t-1} \cap \mathcal{Y}_{t-1}) \tag{3.12}$$

This joint probability can be rewritten as the product of two factors as follows:

$$P(X_t \cap Y_t | \mathcal{X}_{t-1} \cap \mathcal{Y}_{t-1}) = P(Y_t | \mathcal{X}_{t-1} \cap \mathcal{Y}_{t-1}) \cdot P(X_t | Y_t \cap \mathcal{X}_{t-1} \cap \mathcal{Y}_{t-1}) \tag{3.13}$$

Now, we say that *X* does not Granger-cause *Y* if it is true that in the first of the two factors, Y_t*only* depends on its history but not on the history of *X*, that is:

$$P(Y_t | \mathcal{X}_{t-1} \cap \mathcal{Y}_{t-1}) = P(Y_t | \mathcal{Y}_{t-1}). \tag{3.14}$$

We then say that the variable *Y* is independent of the entire history of *X*, that is $\mathcal{X}_{t-1}$, conditional on its history $\mathcal{Y}_{t-1}$:

$$Y_t \perp\!\!\!\perp \mathcal{X}_{t-1} | \mathcal{Y}_{t-1}. \tag{3.15}$$

That is to say, the history of *X* does not convey any relevant information for assessing the effect *Y*. Notice that this characterisation is basically consistent with the main requirements of probabilistic theories of causality as discussed in Chapter 2, in particular with respect to the use of statistical relevance and temporal priority of causes.

There are standard criticisms, though. One concerns the fact that Granger-causality does not enable us to deal with contemporaneous causal links. A second one says that since Granger-causality is typically a bivariate procedure, it is vulnerable to the 'third-variable' problem, namely there might be a third variable that screens-off the correlation between the histories of *X* and *Y*. A third problem concerns the direction of the causal relation: Granger-causality tests may indicate an incorrect causal direction when agents anticipate values of economic variables. Additionally, Granger-causality directly relates to generic causal relations but does

not allow a straightforward interpretation for single-case causal relations. In spite of this, Granger-causality does provide a powerful test for the presence of linear non-contemporaneous causal links.

3.1.4 Rubin's Model

The set up often referred to as Rubin's model is due to work by Donald Rubin (1974, 1978) and by Paul Holland (1986, 1988). The same ideas are also employed in the so-called matching approach (see for instance Rosenbaum and Rubin 1983; Holland and Rubin 1988). Freedman (2005) points out that this type of statistical model for causation traces back to Neyman's work on agricultural experiments in the early twentieth century, therefore he prefers referring to this model as the Neyman-Holland-Rubin model.

The basic idea behind Rubin's model is to measure the effects of the causes rather than to explicate causes of effects. Simply put, consider a large population of individuals; this population is divided into two subpopulations: one in which individuals receive the treatment (treatment group) and one in which individuals do not (control group). Subjects are assigned to the treatment group or to the control group at random. The notion of *potential* plays here a central role: each experimental unit—i.e. each subject in the population—has to be potentially exposable to the treatment, whence the random assignment to the treatment or to the control group. Thus, each individual has two potential responses: the first response is observed if the individual is assigned to the treatment, and the second is observed if the individual is assigned to control. Obviously, both responses cannot be observed simultaneously for the same individual. Causal effects are comparisons among values that would have been observed under all possible assignments of treatments to experimental units. More precisely, according to Rubin (1978: 34–35):

> A causal effect of one treatment relative to another for a particular experimental unit is the difference between the result if the unit had been exposed to the first treatment and the result if, instead, the unit had been exposed to the second treatment.

In other words, we are interested in three parameters:

(i) The average response, if all individuals were assigned to treatment.
(ii) The average response, if all individuals were assigned to control.
(iii) The difference between (i) and (ii).

The third parameter is the *average causal effect* or the average treatment effect. Following Holland's notation (Holland 1988), within a population of units U, S is the variable indicating the causes each unit u in U is exposed to: either treatment (t) or control (c). Y is the response variable and measures the effect of the cause; the two possible responses are Y_t and Y_c. So $Y_t(u)$ indicates the value of the response that would be observed for a given unit u if u were exposed to treatment t. Similarly, $Y_c(u)$ indicates the value of the response that would be observed for a given unit u

if u were exposed to control c. The average causal effect is then defined as $Y_t(u) - Y_c(u)$.

As just pointed out, $Y_t(u)$ and $Y_c(u)$ cannot be observed simultaneously. The average causal effect T of t over U is then defined as the *expected value* of the difference $Y_t(u) - Y_c(u)$ over the u's in U, that is $E(Y_t - Y_c) = T$. In other words, to calculate the average causal effect we are now interested in three estimates:

(i) The average response among individuals assigned to treatment.
(ii) The average response among individuals assigned to control.
(iii) The difference between (i) and (ii).

The third estimate is exactly the quantity $E(Y_t - Y_c) = T$. This can be rewritten as $T = E(Y_t) - E(Y_c)$. This last equation reveals the fact that information on different units *that can be observed* can be used to gain knowledge about T. However, such causal inference about the average causal effect is possible at the price of an untestable assumption of homogeneity of the units in U.

As Freedman (2005) points out, Rubin's model is well suited to experiments or quasi-experiments, where assignments can, at least in principle, be manipulated by the investigator. However, Holland (1988) goes further arguing in favour of the application of this model in other domains such as econometric models for employment discrimination. On the grounds that the two quantities Y_t and Y_c cannot be simultaneously observed, Dawid (2000, 2007) argues against this approach and advocates a Bayesian decision-theoretic framework for causal analysis.

3.1.5 Multilevel Analysis

Recently, a different model—called multilevel analysis or hierarchical modelling—has been developed and adopted in various disciplines, from demography to epidemiology. Multilevel analysis is a methodology for the analysis of data with complex patterns of variability, the underlying assumption being that data shows a hierarchy that cannot be neglected in the analysis. In a less technical vocabulary, the idea is this. Different social sciences study human behaviours and phenomena and try to find and explain (causal) mechanisms responsible for them. However, the object of the discipline does not directly specify the level of aggregation at which analyses have to be run. For instance, economics is interested in the production, distribution and consumption of wealth; however, there is no a priori specification of whether analyses have to concern individuals, markets, firms, or nations. Likewise, demography has no prior specification of whether the study is at the level of family, local population, or national population.

Thus, multilevel analysis recognises the existence of a multiplicity of levels, and tries—within the framework of a *single* model—to specify the relations holding among individuals and/or among different levels of aggregation. In other words, this approach recognises that the grouping of individuals introduces an influence of the group on its members, and, conversely, that members have an influence on

the group's behaviour. Failure to recognise this twofold source of influence and variability may lead to two types of fallacy: the atomistic fallacy and the ecological fallacy.

Firstly pointed out in a 1950 paper by Robinson, the ecological fallacy consists of inferring erroneous individual behaviours from aggregate measures. Robinson pointed out, for instance, that correlations between two characteristics measured on a binary basis among individuals (e.g. being black and illiterate in the US), or by proportions in regions (e.g. proportions of black and illiterate people in the population) were generally not identical and could even carry opposite signs. Conversely, the atomistic fallacy arises when, analysing individual behaviours, the context in which such behaviours occur is neglected. I shall get back to these fallacies in more detail in Chapter 6.

The solution is then a model in which different levels can simultaneously be included. Very good introductions to multilevel modelling are those of Goldstein (1987, 2003), Snijders and Bosker (2004), and Courgeau (2003, 2004a, 2007a) in which epistemological and methodological problems are also discussed at length. For convenience of exposition, this section will mainly go along with Snijders and Bosker's textbook. Instead, other epistemological and methodological related aspects will be discussed in Chapter 6, following up some of Courgeau's arguments.

The first thing to clarify is the type of claims made in (i) an individual-level model, (ii) an aggregate-level model, and (iii) a multilevel model. Let us consider again the case study on migratory behaviour presented in Chapter 1. Individual-level models explain individual-level outcomes by individual-level explanatory variables—for instance, we might explain the individual probability of migrating through the individual characteristics of being/not being a farmer. Aggregate models explain aggregate-level outcomes through explanatory aggregate-level variables—for instance, we might explain the percentage of migrants in a region through the percentage of people in the population having a certain occupational status (e.g. being a farmer). Finally, multilevel models make claims across the levels—from the aggregate-level to the individual-level and vice-versa; for instance, the multilevel model of migration in Norway explains the individual probability of migrating for non-farmers through the percentage of farmers in the same region: although being a farmer in a region lowers the chances of migration, as the percentage of farmers increases, the probability of migration raises in the whole population because non-farmers tend to migrate more.

Let us now look at the characteristics of multilevel models. Snijders and Bosker (2004: 1) identify in the contextual analysis and in the mixed models effects the two tributaries of multilevel analysis. On the one hand, contextual analysis has focused on the effects of the social context on individual behaviour; on the other hand, mixed effect models assume that some of the coefficients are fixed and others are random. Let us see how the two are formally combined in a single model. Let us follow again Snijders and Bosker (2004: 38) who describe a hierarchical linear model as "a type of regression model that is particularly suitable for multilevel data". Before seeing what makes hierarchical models suitable for multilevel data, let us fix the notation first.

For the sake of simplicity, we shall deal with a model in which only two levels are present. Thus, level-one units refer to 'individuals' and level-two units refer to 'groups'. The number of groups in the data is denoted by N; the number of individuals in the groups may vary from group to group and is denoted by n_j for the group $j(j = 1,2,\ldots,N)$. The total number of individuals is denoted by $M = \sum_j n_j$. The main difference between a regression model and a hierarchical model is that the latter contains more than one error term, one for each level of analysis. Also, hierarchical models normally take the dependent or response variable to be at level-one—i.e. explanatory variables at level-one or above are used to account for what happens at the lowest level. Again, for the sake of simplicity, let us assume that we have at our disposal one explanatory variable at each level. Let us denote with j the index for the groups ($j = 1,\ldots,N$) and with i the individuals within the groups ($i = 1,\ldots,n_j$). The indices can be considered as case numbers and the numbering for individuals starts again in every group. For individual i in group j, we have the following variables:

(i) Y_{ij} is the dependent variable.
(ii) X_{ij} is the explanatory variable at the individual level.

for group j we have:

(iii) Z_j is the explanatory variable at the group level.

It is worth noting that indices i and j precisely indicate what the variables depend on. In this way multilevel modelling conveys the idea that the response variable Y bears an individual aspect as well as a group aspect. With these variables at hand, a classical multiple regression model could be written as follows:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_j + \varepsilon_{ij} \tag{3.16}$$

or, introducing an interaction effect between Z and X:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_j + \beta_3 z_j x_{ij} + \varepsilon_{ij}. \tag{3.17}$$

These models suggest that the response variable Y is fully explained by the group variable Z and the individual variable X. However, these models cannot make a difference whether individuals belong to the same or to different groups. So, in order to better represent the nested structure of data, we have to let the regression coefficient vary from group to group, formally:

$$Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \beta_2 z_j + \varepsilon_{ij}. \tag{3.18}$$

Now the regression coefficients β_{0j} and β_{1j} can vary according to the group. However, if the coefficients β_{0j} and β_{1j} are kept constant—i.e. do not depend on j—then the last model simply gets us back to the previous one in which the nesting structure has no effect. Instead, if those coefficients *do* depend on the group, we have to take into account how the nesting structure influences the effects of X and Z on Y. This is done by allowing the intercept β_{0j} and the slope β_{1j} to vary randomly.

In plain words, considering the classical example in the educational field that analyses the nesting structure of pupils within classrooms, this would be tantamount to saying that the effect of socio-economic status of pupils on their scholastic achievement also varies according to the level-two aggregation, i.e. the classroom.

Multilevel models also rest on specific assumptions. These assumptions concern:

(i) Independence and distribution of errors:

(a) Errors are independent at each level and between levels.
(b) At level-one errors are normally distributed with 0 mean and variance σ^2.
(c) At level-two errors follow the multivariate normal distribution with constant variance matrix.

(ii) Specification of the variables having random slopes implying a certain variance and correlation structure for the observations.
(iii) For multilevel linear models, the linear dependence of the dependent variable Y on the explanatory variables and on the random effects.

3.1.6 Contingency Tables

When variables involved in the analysis are non-metric, categorical data analysis (CDA) is more often employed instead. CDA has a very long history. It began in the early 1900s, when Karl Pearson and Udny Yule were debating on measures of associations, and two decades later CDA took advantage of significant contributions by Sir Ronald Fisher. The first lucid exposition of the use of contingency tables in sociology is due to Boudon and Lazarfeld in *L'analyse empirique de la causalité* (1966). In recent years, clear presentations and further improvements of CDA are available in the works of Hellevik (1984) or Agresti (1996). I shall first state what categorical data is and then present how data is analysed by contingency tables.

A categorical variable is one for which the measurement scale consists of a set of categories. Categorical scales are very often used in the social sciences to measure attitudes and opinions on several issues and they are also used in the health sciences. Among categorical variables two types of measurement scales can be distinguished. When variables have ordered scales they are called ordinal—e.g. measuring attitudes towards legalization of abortion, the scale might be 'disapprove in all cases', 'approve only in certain cases', 'approve in all cases'. Instead, when variables have unordered scales they are called nominal—e.g. religious affiliation might involve the following categorization: Catholic, Protestant, Jewish, other. For nominal variables, the ordering listing is clearly irrelevant, and the statistical analysis will not depend on that ordering, whereas methods designed for ordinal variables will in general take the ordering into account.

Categorical data consist of frequency counts of observations occurring in the response categories. Consider the simplest case, where only two variables X and Y are involved: X is said to have i levels, and Y is said to have j levels, according to the number of categories that X and Y involve. The ij possible combinations of

Table 3.1 Frequencies of sentence according to social status, criminal record, and severity of sentence

Data Matrix			Frequency
Status	Record	Sentence	–
High	Criminal	Severe	25
High	Criminal	Lenient	15
High	Clear	Severe	2
High	Clear	Lenient	22
Low	Criminal	Severe	95
Low	Criminal	Lenient	22
Low	Clear	Severe	17
Low	Clear	Lenient	18
Total	–	–	216

outcomes are then displayed in a rectangular table having i rows and j columns. The cells of the table in fact represent the ij possible outcomes and contain frequency counts of outcomes. Tables thus construed are called contingency tables or crosstabulations. Such ordinary percentage tables can be analysed by adopting an explicit causal framework.

Consider the following example (Hellevik 1984). A study has been carried out about decisions made in the Norwegian courts. A sample of 216 men convicted of theft is analysed, taking the following variables into account: social status (based on occupation and income), criminal record (previously convicted or not), and severity of the sentence received. For the sake of space, suppose we do not list every single defendant separately, but contract the matrix by grouping together units with similar value combinations as in Table 3.1 (Hellevik 1984).

Suppose now we want to summarise information about the variable 'sentence' as shown in Table 3.2. The second column shows the absolute frequency for the two values of the sentence variable. The same information can be expressed as a percentage (third column) by dividing each of the frequencies by their sum and multiplying by 100. If we don't multiply by 100 we obtain the proportion (forth column). It is worth pointing out that while the absolute frequency conveys very little information, the proportion is much more informative, for we automatically know that just below two-thirds of all defendants were sentenced to imprisonment.

Table 3.2 Summary information about the variable sentence

Sentence	Absolute Frequency	Percentages	Proportions
Severe	139	64	0.64
Lenient	77	36	0.36
Total	216	100	1.00

Table 3.3 Trivariate table with sentence as the dependent variable and social status and record as independent variables

	Absolute Frequency				Proportions			
Social Status	High		Low		High		Low	
Record	Criminal	Clean	Criminal	Clean	Criminal	Clean	Criminal	Clean
Sentence					–			
Severe	25	2	95	17	0.63	0.08	0.081	0.49
Lenient	15	22	22	18	0.37	0.92	0.09	0.51
Sum	40	24	117	35	1.00	1.00	1.00	1.00

Proportions also have an interesting interpretation in terms of probability. A proportion may be thought of as the probability that a unit randomly drawn from the data matrix will have the value in question. That is, if we randomly draw a unit from the matrix, the probability that a given defendant will have received a severe sentence is defined as the long-run proportion of times that this result occurs.

Consider the simple case in which we analyse a bivariate crosstabulation; for instance, we take into account the sentence and the social status variables. Such a table will tell us what proportion of offenders with low status was sent to prison, compared to high status offenders. Computing the difference between the proportion having a certain value of the dependent variable within each of the two groups defined by the dependent variable, we will get the degree of *statistical association* between those two variables. If the difference in proportion is zero, this will be taken as no association.

However, for the purpose of causal analysis, we will usually be more interested in analysing *multivariate* crosstabulations, i.e. tables in which information about more than two variables is displayed. In building those multivariate crosstabulations, we explicitly choose one variable as being the dependent variable and the others as being the independent ones.

Consider again the study on decisions made in the Norwegian court. A trivariate table would turn out to be useful were we to suspect that the difference between high- and low-status defendants with regard to punishment is related to differences between the two groups with respect to their prior conflicts with the law (either clean or criminal). Table 3.3 displays ‘sentence’ as the dependent variable, and ‘social status’ and ‘record’ as the independent variables.

This trivariate table allows us to see that there are indeed differences in the probability of a severe sentence. In particular, low-status defendants with a criminal past have a proportion of 0.81 to be sent to jail. Conversely, high-status defendants with no criminal past have a low proportion (0.08) sent to jail.

The question is now, as Hellevik says, to determine “what each of the independent variables contributes to the variation between the groups”. In a nutshell, the idea is to calculate two differences in proportions, namely between high- and low-status defendants who have a clean record, and similarly for defendants with a criminal record. The variable ‘record’ will now play the role of a *control variable*, that is its

value is kept constant when the association between status and sentence is calculated (see Hellevik 1984 for details).

Needless to say, analysis by contingency tables requires a structured theoretical framework for the *causal* interpretation to be justified. I shall leave a thorough discussion of this issue aside and just mention that the choice of variables, background knowledge, confounding factors and other metaphysical assumptions play a fundamental role. In contingency tables differences in proportions play the analogue of regression coefficients in SEM, giving highly similar results. Thus, to some extent, the causal framework for contingency tables rests on the same features as SEM, namely on background knowledge, choice of variables, issues of confounding and control, etc. Those topics will be discussed again at length in section 4.3.

3.2 Hypothetico-Deductive Methodology

Hypothetico-deductivism is the view according to which, roughly speaking, the scientist first formulates a hypothesis and then tests it by seeing whether the consequences derived from the hypothesis obtain or not. Famously, Karl Popper held that if the predicted consequences actually obtain, nothing can be said about the hypothesis because science is about *falsifying* rather than confirming: a single positive instance cannot confirm the hypothesis as there still is the possibility that it will be contradicted by evidence in the future. This extreme position was mitigated later on (Popper himself came to accept the notion of 'corroboration') and it is widely agreed that positive instances do confirm, at least to some extent, scientific hypotheses. Conversely, if the predicted consequences do not obtain, the hypothesis has to be rejected and a completely new one has to be formulated.

The hypothetico-deductive methodology is an important feature of the vast majority of causal models used in contemporary social science. Model building and model testing essentially involve three stages:

1. Formulate the causal hypothesis.
2. Build the statistical model.
3. Draw consequences to conclude to the empirical validity or invalidity of the causal hypothesis.

As we saw in Chapter 1, social scientists start with the formulation of a causal hypothesis. Such a hypothesis, however, does not come from a *tabula rasa*, but emerges within a causal context, namely from background theories, from knowledge concerning the phenomenon at issue, and from a preliminary analysis of data. The causal hypothesis, which is also called the 'conceptual hypothesis', is not analysable *a priori*, however. That is to say, its validity is not testable by a logico-linguistic analysis of concepts involved therein. On the contrary, testing the validity of the causal hypothesis requires building a statistical model, and then drawing consequences from the hypothesis. For instance, in the case study presented in section 3.1.4, Adam et al. formulate two hypotheses to test: (i) no causal link from

health to wealth, and (ii) no causal link from wealth to health. On the basis of background knowledge and previous studies researchers expect the first hypothesis to be accepted and the second to be rejected. The consequences they draw concern exactly what they expect to be accepted and rejected during the testing stage. The estimation of the statistical model and hypothesis testing will allow us to conclude to the *empirical* validity or invalidity of the causal hypothesis. If the model is correctly estimated and fits the data, the hypothesised causal link is accepted, rejected otherwise. The hypothetico-deductive structure of causal modelling is thus apparent: a causal hypothesis is first formulated and *then* put forward for empirical testing, i.e. the causal hypothesis is confirmed or disconfirmed. That it to say, the causal hypothesis is *not directly inferred* from the data gathered, as is the case with inductive strategies, but accepted or rejected depending on the results of tests. For a lucid account of hypothetico-deductive and inductive methodology see also Williamson (2005a: ch. 8).

Perhaps the label hypothetico-*deductivism* is somehow misleading because, strictly speaking, there is no deduction going on. In causal modelling, hypothetico-deductivism does not involve deductions *strictu sensu*, but involves a weaker inferential step of 'drawing consequences' from the hypothesis. That is to say, once the causal hypothesis is formulated out of the observation of meaningful co-variations between the putative cause and the putative effect and out of background knowledge, we do not require data to be *implied* by the hypothesis, but just that data conform to it. Here, 'conform' means that the selected indicators *adequately* represent the conceptual variables appearing in the causal hypothesis. Thus, this way of validating the causal hypothesis is not, strictly speaking, a matter of deduction, but surely is, broadly speaking, a deductive procedure. More precisely, it is a *hypothetico*-deductive procedure insofar as it goes in the opposite direction to inductive methodologies: not from rough data to theory, but from theories to data, so to speak. For a discussion of an example of the H-D method at work in the social sciences, see also Cartwright (2007a: ch. 2). Cartwright, as many others both in the philosophical and scientific literature, calls the methodology of causal models hypothetico-deductive but she also warns us about the weaker form of deductivism hereby involved.

In the presentation of covariance structure models, Aish-Van Vaerenberg (1994) stresses that CSM employ a confirmatory logic (*logique confirmatoire*). The validity of the hypothesised causal structure is stated thanks to (i) formal tests of adequacy between the model and empirical data, and (ii) statistical significance of coefficients, provided that the model is uniquely identified. Also, CSM rely on an *epistemological* rather than ontological conception of causation: the problem of causality is located at the level of *knowledge*. Thus, a causal theory is a reconstruction of reality in terms of causes and effects, according to *knowledge* we can attain about them. To sum up, CSM are a tool that allows us to test the empirical validity of a causal structure, on the basis of a statistical verification of a mathematical model.

As I understand this argument, Aish does not claim that a particular *logic of confirmation* is used; rather, that CSM employ a hypothetico-deductive methodology

to *confirm* (or disconfirm) a given causal hypothesis. Differently put, philosophically speaking, the terminology used by Aish-Van Vaerenberg is rather misleading, although her ideas seem correct. This last remark leaves room for a couple of refinements in the H-D methodology.

To begin with, let me clarify one issue. There is a sharp difference between hypothetico-deductive or inductive methods on the one hand, and deductive or inductive inferences, on the other. The former are methodologies for the testing or for formulating hypotheses, whereas the latter are types of inference. On the one hand, H-D methods *confirm* (or disconfirm) hypotheses, while inductive methods are employed to *discover* hypotheses. On the other hand, deduction is a *non-ampliative inference* from what is known to what is known, whereas induction is an *ampliative* inference from what is known to what is *not* known yet. Let me spend a few more words on this issue, for the distinction might be self-evident to logicians or philosophers of science, but it is far from being clear in the scientific literature.

For instance, the *Journal of Epidemiology and Community Health* published a glossary concerning causality in the public health sciences (Susser 2001). We find here (Susser 2001: 376):

> *Deductive logic* predicts particular outcomes from prior general hypothesis—that is, it proceeds from the general to the particular.

In valid deductive arguments the premises *logically entail* the conclusion, that is to say, given the truth of the premises it cannot be the case that the conclusion be false. Thus the matter is not going from the general to the particular. The premises logically entail the conclusion because what the conclusion states is already contained, either explicitly or implicitly, in the premises. When people say that deduction goes from the general to the particular, they typically have in mind a syllogistic form such as:

> All men are mortal;
>
> Socrates is a man;
>
> Therefore, Socrates is mortal.

However, we can safely use modus ponens—a valid inference rule of deductive logic—to make inferences *from the particular to the particular*. For instance, from the premises:

> If Socrates is a man, then Socrates is mortal;
>
> Socrates is a man;

we can deductively infer by *modus ponens* that Socrates is mortal.

Let us now read the entry *inductive logic* (Susser 2001: 376):

> Inductive logic seeks to generalise by reasoning from assembled particular observations […].

This is only partly correct. Inductive logic deals with inferences that are less-than-certain. A good inductive argument is one in which the premises provide a good

degree of support to the conclusion. Very often probability theory is applied to sentences to represent a measure of the degree of support. The correct part is that we might want to generalise from a given number of observations: we observe several ravens, one after another, and they are all black. We then generalise and say that *all ravens are black*. The trouble is that we cannot observe *every* raven in order to confirm the hypothesised generalisation. This is the well-known raven paradox discussed by Hempel (1965).[2] However, not all inductive inferences go from the particular to the general. For instance, after having observed ten ravens one after the other, and having observed that all are black, I might predict that *the next raven* will be black. This is again an uncertain inference but it is not a generalisation—it is an inference to the next case, i.e. a prediction, and it is inductive.

Let us now focus on hypothetico-deductive methodologies. H-D do not require us to adopt a *strict* covering-law model, nor any *particular* measure of confirmation. The covering-law (C-L) model has been strongly criticized for several reasons. For instance, the concept of *law* seems to be too demanding. Nevertheless, the *deductive character* of C-L models appears to be the most problematic aspect. However, it might be argued that not all C-L models are deductive in character, as the deductive-nomological (D-N) is; in fact, in his seminal 1948 paper, Hempel himself accounted for *inductive-statistical* models as well. Moreover, even the statistical-relevance (S-R) model developed by Salmon (1971, 1984) belongs to the C-L models, since it employs *generalisations* in an essential manner. Indeed, C-L models have been redeemed in recent works by Little (1993, 1995b) and Irzik and Meyer (1987), among others. In sum, H-D strategies do not commit us to a *strict* C-L model as the D-N one, for the covering-law model is quite general too.

As far as the concept of law is concerned, a weaker concept of *empirical generalisation* will do. Such a concept of empirical generalisation seems to be suitable because universal statements that enter causal models are not laws of nature *strictu sensu*. Rather, they are *statistical* statements concerning the population under analysis. Indeed, those statistical generalisations surely lack the necessity which is proper to *laws* of nature. If there are any. In fact, the very concept of law of nature has been widely criticized in the literature. Recent accounts (e.g. van Fraassen (1989) and Giere (1999)) try to show that science does not need universal and necessary laws. Others even challenge necessity, by saying that some contingently true propositions are—or could be—laws of nature since laws are just singular statements about universals. It is worth noting that even though the very concept of law is challenged, this does not rule out H-D methodology altogether: causal hypotheses are confirmed or disconfirmed depending on whether the consequences drawn from general statements—were they laws of nature *strictu sensu*, contingent laws of nature, or humble empirical generalisations—fit the empirical data at hand.

[2] The paradox goes even further. 'All ravens are black' has the logical form of the conditional $A \rightarrow B$ which in this case reads 'if a given object is a raven, then it is black'. This conditional, in turn, is equivalent to its contrapositive, i.e. $\neg B \rightarrow \neg A$, which in plain English reads 'if something is non-black, then it is a non-raven', that is 'all non-black things are non-raven'. Consequently, even a white shoe would eventually confirm 'all ravens are black'.

So far so good. Thanks to this H-D methodology, the hypothesised causal relation is confirmed (or disconfirmed). Does this commit us to any particular measure of confirmation? Not quite, for we might say that causal hypotheses are confirmed to the extent to which the model fits the data. This is just a *qualitative* claim, and different measures of confirmation may fulfil this condition, although they might give different *quantitative* results. Williamson (2005a: § 8.2) makes a similar point on the generality of the H-D scheme).

The difference between hypothetico-deductive and inductive methodologies emerges even more clearly when we compare ideas behind exploratory data analysis (EDA) and confirmatory data analysis (CDA). EDA is mainly concerned with the encouragement of hypothesis formulation, whereas CDA deals with experimental design, significance testing, estimation and prediction, where the *prior* formulation of hypotheses is needed. I have said enough about the philosophy of confirmatory statistics discussing the H-D methodology, so I will now say a few more words on the philosophy of EDA.

As a matter of fact, there has not been much discussion in the literature specifically about the philosophy of EDA and the philosophy of CDA. These are coinages due to I. J. Good, who published the paper "The philosophy of exploratory data analysis" in 1983. As far as I know, not much reaction followed Good's paper, except a reply in 1985 by Stanley Mulaik published in the same journal.

In this paper, Good (1983b) points out that EDA goes beyond descriptive statistics, in the sense that the aim is to 'look at data' in order to recognize *patterns* in the data. Techniques of descriptive statistics are designed to match the salient features of data; instead, this 'looking for patterns' in EDA somehow recalls the Baconian method of science. As is well known, the Baconian inductive method starts from sensible experience and moves up to the formulation of axioms or laws of nature. Such laws are derived from the tables of presence and absence, and from tables of comparison and degrees. This analysis of observations leads to the *discovery* of scientific truths, provided that fallacious reasoning—the *idola*—are avoided.

Hence, EDA is a modern form of the Baconian method, because causal hypotheses are formulated only after having observed the data. Techniques performed by Glymour and his associates, or by Pearl, are *inductive* strategies for the discovery of causal relations precisely in this Baconian sense: algorithms are designed to infer causal relations directly from the data gathered, independently of any prior conceptual hypotheses. As Mulaik (1985: 410) puts it:

> For Good, the relevance of the Baconian method to prevailing statistical practice is in the emphasis on first looking at the data *before* formulating hypotheses. (My emphasis.)

Let me emphasize that my aim here is not to appraise H-D methodologies over inductive ones, but just to point out that much of contemporary social science takes advantage of the hypothetico-deductive methodology, as we saw earlier in the analysis of structural models, covariance structure models, Granger-causality, etc.

Before closing this section, let me linger a bit on the philosophy of exploratory statistics. As I mentioned, Mulaik echoes Good's paper not to challenge his claims, but rather to clarify some issues in the history of statistics. In describing exploratory data analysis as a hypothesis-*formulation* technique, Good concedes the importance of this methodology for the subsequent stage of hypothesis-*testing*. What Mulaik contends is that, *historically*, the modern forms of exploratory statistics did not develop in connection with the hypothesis-testing framework. His explanation for this weak connection is that exploratory statistics developed within the context of an empiricist philosophy of science "that emphasised description, sticking to the 'fact', induction from given facts, and (sometimes) hostility towards hypotheses, especially those about the unseen" (Mulaik 1985: 411).

I strongly recommend this article for the reader interested in statistical thinking. Mulaik brings us from mid-eighteenth century to the present-day, and brightly illustrates how the philosophies behind the works of Laplace, Whewell, Mill, Quetelet, Galton, Pearson, Yule, Kendall, etc. have evolved. Nonetheless, by mentioning Mulaik's contribution I do not intend to make an historical point, but a *theoretical* one. After all, Mulaik raises a series of issues of primary importance. Should we be afraid of metaphysical hypotheses and assumptions about causal relations? Is a realist position compatible with causal modelling? Does a Bayesian framework necessarily lead to an antirealist position about causation? Unfortunately, to fully answer those questions falls far beyond the scope of this book. However, from time to time these issues will arise again.

3.3 Difficulties and Weaknesses of Causal Modelling

Since the first developments of quantitative causal analysis, both practising and methodologically-oriented scientists discussed various issues that make causal research a complex and challenging enterprise. However, those discussions have been led in a rather disperse manner—no systematic exposition has been offered so far. In the following, I will attempt an overview of the difficulties and weaknesses of causal modelling by choosing a particular strategy of exposition. Following the hypothetico-deductive methodology discussed in the previous section, we shall see which threats causal analysis faces at each step. A significant effort will be made in stating the sources and the consequences of the problem.

By way of reminder, the H-D methodology involves three subsequent stages:

1. Formulate causal hypothesis.
2. Build the statistical model.
3. Draw consequences to conclude to the empirical validity/invalidity of the causal hypothesis.

As we have seen, the process of model building involves a continuous interaction between background knowledge and a sequence of statistical procedures for elaborating and testing hypotheses. Prior theorising of out-of-sample information includes,

in particular, the selection of variables deemed to be of interest and the formulation of a causal hypothesis (also called the conceptual hypothesis). Then, iteratively, we build up the statistical model and we test the adequacy between the model and the data to assess the empirical validity or non validity of the causal hypothesis.

3.3.1 Background Knowledge and the Formulation of Causal Hypotheses

To the practising scientist, the paragraph above might look somehow obvious. *Of course* model building involves a continuous interaction between background knowledge and statistical procedures, and *of course* the formulation of the causal hypothesis is included in the theorising out-of-sample information. Not *so* obvious, though. On the one hand, it is far from being clear what exactly background knowledge is or what it ought to include, and, on the other hand, how such knowledge ought to be formalised into a causal model.

Hypothetico-deductive causal models are primarily concerned with this problem. In fact, inductive methodologies, such as Spirtes et al. (1993), attempt to infer causal relationships merely from associations found in observational data, *without* appealing to any background knowledge. This is indeed a controversial and contested result; however, our concern here is not to assess one methodology over another, but rather to understand why background knowledge, although necessary, is a weak point of causal modelling.

Some, for instance Meek (1995) and Williamson (2005a), suggest that background knowledge includes causal constraints—e.g., prior knowledge of existence of causal relationships or of temporal ordering of variables—and probabilistic constraints—e.g. information about dependencies between variables. Halpern and Pearl (2005a, b) give more hints about knowledge to use in causal models. According to them the structural modelling approach is a "formal representation of causal knowledge and a principled way of determining actual causes from such knowledge" (Halpern and Pearl 2005a: 877). In their approach, background knowledge encoded in the model includes implicit assumptions about the causal scenarios, the causal mechanism (changing the causal mechanism would result in changing, accordingly, the system of equations), and the decision about the set of variables used and depending on the causal role they play (dependent or independent). They acknowledge the subjectivity of the choice of the model and the fact that such choice also depends on what the model is being used for.

However, what is still lacking is a neater characterisation of the causal context, which is the source of the causal hypothesis. As we shall see later in section 4.3, the specification of the causal context is of primary importance, because if we content ourselves with the features just mentioned, given an appropriate statistical dependence, we might accept nonsense causal relations such as the increasing number of storks in Alsace causing an increasing number of births in the same region. Surely this is a kind of causal claim that we would like to exclude straight away. But if

we fail to specify *where* the causal hypothesis comes from, we have no means of excluding such causal claims *a priori*, and we are unable to *theoretically* justify the causal hypothesis.

3.3.2 Theorising or Modelling?

A possibility is that causal hypotheses come from something more specific than the causal context, namely available *theories*. For instance, a causal hypothesis concerning migration behaviour in the Norwegian population (see the case study presented in section 1.5), could be formulated from general demographic theories of migration. Such a position hides at least three assumptions: (i) we have a clear-cut concept of theory; (ii) the theory 'dictates' the causal hypotheses to test; and (iii) general theories are generated 'outside' the causal model that is specifically designed to test the causal hypothesis. Each of these assumptions can be challenged, though.

Hubert Gérard (1989, 2006) is an excellent example of the difficulty of reaching a consensus as to what constitutes a theory in the social domain. Gérard tackles the issue of theorising in demography and, more generally, in the social sciences. According to him, one possibility is to walk a normative path by imposing a definition of what a theory is and to analyse its appropriateness and feasibility. A second possibility is to walk an empirical path by examining what is actually called 'theory' in the social sciences. This second route—claims Gérard—is preferable as it avoids two extreme positions: on the one hand, a theory in social science has to meet the requirements of the hard sciences and, on the other, anything is a theory.

Gérard sees theorising as a demarche leading to the construction of a theory. In his words (Gérard 1989: 271) theorising is a

> […] process of critical systematisation of the *acquired knowledge*, in a specified body of propositions that entail hypotheses liable to be tested against empirical data. (My translation[3] and emphasis.)

By 'acquired knowledge' Gérard means previous theories as well as any kind of background knowledge, including researchers' prejudices. This process, as Gérard correctly points out, is far from being linear, and at each stage we encounter conceptual and practical difficulties. The process of theorisation, in his account, articulates on four *axes*:

1. The scope of the theory under construction, that can be limited to the specific cases at hand or can aim at some sort of universality.
2. The formulation of the theory, involving precise concepts and relations among them.

[3] In the original: "[…] processus de systématisation critique des connaissances acquises, dans un corps désigné de propositions qualifiées engendrant des hypothèses susceptibles d'être soumises à l'empirie".

3. Juxtaposition of propositions and their combination in a deductive or, more often, hypothetico-deductive set up.
4. Confrontation with empirical evidence.

Following Blalock (1968a, b) and Duchêne and Wunsch (1985), Gérard then distinguishes six *steps* in the process of theorising:

1. Research question.
2. Principal theory.
3. Auxiliary theory.
4. Mode of confrontation with empirical evidence.
5. Techniques used to confront with empirical evidence.
6. Results.

The research question sets up the object of study, and although it conceptually precedes all the other stages, it is of course highly influenced by them. The principal theory is usually presented in natural language and helps in constructing concepts and relations among concepts in a clear and non ambiguous way. These concepts and their relations are then translated in terms of indicators, according to the mode of confrontation and the techniques—i.e. the causal model—chosen for the case at stake. This adaptation and translation of the principal theory leads to one or several auxiliary theories. The principal and auxiliary theory, in turn, lead to the formulation of hypotheses to be put forward for empirical testing. Results of tests will depend on the chosen model (e.g. quantitative or qualitative methods) but also on the available data. By evaluating results we will be able (i) to assess the adequacy of the principal theory as an answer to the research question, and (ii) to consider how to orient further research.

It is worth noting that this characterisation of theorising is wholly equivalent to the hypothetico-deductive methodology of causal models discussed in section 3.2. Thus, a fundamental difficulty causal modelling faces concerns its own status, which leads to new questions. Are modelling and theorising the same thing? If so, does it mean that a good causal model is ipso facto a good theory? But if this is true, why do the social sciences oft complain of not having *theory*? To provide a coherent account of *theory* unfortunately falls beyond the scope of the present work. However, in section 6.1, I will advance the view that causal modelling is the modelling of mechanisms, where the mechanism is not just an assembly of (physical) gears, but can be represented as an arrangement of variables requiring a certain *theoretical* plausibility. This is a first step towards a broader and more complete concept of causal model that goes beyond mere statistical testing of hypotheses. But let us now develop more on the problem of lack of theory.

3.3.3 Lack of Theory

The insufficient specification of the causal context has also another important consequence: in order to formulate the causal hypothesis, we have to appeal to available

theories. But, as just mentioned, lack of theory is one of the most challenging problems in the social sciences.

The issue has been lucidly tackled, for instance, by Guillaume Wunsch (1995), who provocatively entitled one of his papers "God gave the easy problem to the physicists". The reason why physicists have an easier task, says Wunsch, is that whilst the natural sciences deal with 'things', the social sciences deal with people, who have beliefs, desires, intentions, goals and who, I'd add, are so different from one place to another and from one time to another—physical objects do not display such a great variability. This makes theories in the social sciences, if there are any, extremely local in scope and dependent on time–space; consequently, accumulation of knowledge eventually leading to theory is very difficult to achieve.

Wunsch also discerns other related difficulties. Although in the last few decades numerous conceptual frameworks have been developed, a central paradigm of explanation still seems to be lacking in many disciplines. Often, explanatory frameworks are merely sketched with the consequence that concepts are not clearly defined, the indicators used are not properly justified, and the relations between latent variables are not based on sound theoretical grounds and on prior research findings. Some theories are developed at the macro (i.e. societal) level and others at the micro (i.e. individual) level. Combining both types of results is far from being straightforward: theories developed at one level only may eventually turn out to be inadequate at the other level, as the ecological and atomistic fallacies show (on this point, see later section 6.3). However, theory is indeed needed because it helps in deciding between alternative explanatory structures, or in deciding whether or not to control for some covariates.

Much of what Wunsch says seems to suggest that causal modelling, in order to be of good quality, has to rely on *theory* but he does not tell us *where* theory comes from. If it comes from modelling itself, as Gérard suggested, then it might be the case that we get into a vicious circle. But what would a theory *in the social sciences* state? Isn't this indictment of lack of theory dependent on a meaning of theory that fits the natural sciences and on a concept of law that is exceptionless and universal? This is the topic I shall deal with next.

3.3.4 Establishing Social Laws

Traditionally, the fact that the social sciences cannot establish laws as physics does has been taken as a structural deficiency of the social sciences, the consequence of which was to cast doubt on their very scientificity and rigour. Scientific understanding is often perceived as being objective only if results of research can be treated as 'facts', 'data', that we can generalise into universal laws. This paradigm, as Montuschi (2003, 2006) explains brightly, straightforwardly applies to the natural sciences, but it is less than obvious in social research. Thus the possibility of, and the success in, establishing universal laws has long been considered the hallmark of every *scientific* discipline. However, the adequacy of the positivist paradigm in

social contexts is altogether a controversial issue. In fact, it relies on a standard of objectivity and of scientific practice that well fits the natural sciences, but not the social sciences.

However, this is an extremely complex and vast philosophical debate, of which the problem of establishing social laws only represents the edge of the iceberg. Philosophers, for instance, engage in such quarrels supporting at times very opposite views. As an example, in the recent book *Contemporary debates in philosophy of science*, Hitchcock hosts two rival contributions. Roberts (2004) defends the view that laws are universal regularities and therefore the absence of laws in the social sciences at least establishes a neat difference between them and the natural sciences. Here, needless to say, we are completely inside the received paradigm mentioned above. This narrow view on what a scientific law is traces us back to the attempt to draw a sharp line between the social and the natural sciences— this line being 'objectivity' and 'scientificity'. On the other hand, Kincaid (2004, but see also 1990) argues against a concept of law as universal generalisation. Instead, laws identify (causal) mechanisms and therefore the social sciences do have laws (e.g. the economic law of supply and demand) that work exactly in the same way as the laws of physics. Under Kinkaid's account the social sciences do not suffer from an inferiority complex, so to speak. Nonetheless, we are here embedded in another dominant paradigm—the paradigm of regularity.

The trouble with laws is that they are supported by the received view, according to which regularity is the key notion for causality. In section 4.1, I will undermine the plausibility of the received view. The rationale of causality I propose, based on the notion of variation, will not only shed light on the scheme of reasoning governing causal modelling, but will also contribute, as I will further discuss in section 6.1, to liberate the social sciences from cumbersome 'laws' in favour of nimbler empirical generalisations that describe *variational* statements rather than regular and universal ones.

3.3.5 Concepts and Indicators

Causal modelling aims at detecting and establishing—confirming, according to the H-D methodology discussed in section 3.2—causal relations. I said, causal *relations*. This suggests that causal claims have specific *relata*—the terms, objects, variables that take part in the relationship. What we choose as relata is at least as important as the justification of the presence of a causal relation between them. In causal modelling, these relata are typically concepts represented by variables. Variables, however, can be of very different sorts and the reliability of results depends, among other things, on which variables we choose and on *how* we choose them. In section 6.3, I will offer a taxonomy of the types of variables used in causal modelling and explain the kind of information they convey. In the following, I would like to draw the reader's attention to the difficulty of choosing concepts and/or their

indicators. The issue is of course not new in the social sciences literature (see for instance Zeller and Carmines 1980; Land 1983; Reiss 2001).

In the economic domain, Land (1983) distinguishes three classes of indicators:

(i) Normative welfare indicators.
(ii) Satisfaction indicators.
(iii) Descriptive social indicators.

Their interpretation and use is crucial for two distinct but nonetheless related purposes: policy making and social reporting. The use of social indicators for policy analysis, in particular, requires that analysts agree on three issues: (i) about what needs improving, (ii) that it is possible to decide unambiguously what 'getting better' means, and (iii) that a high degree of aggregation in the indicators exists to facilitate national-level analysis. These three requirements are, as Land (1983: 10) correctly points out, dependent on implicit presuppositions concerning the definitions of normative welfare indicators, of global quality of life and on the identification of key variables. Differently put, indicators are, in the philosophical jargon, *theory laden*. The theory-ladenness thesis (Hanson 1958) taught us that there isn't such a thing as rough data. All observations depend, at least partly, on the theoretical background that we carry with us. Thus the possibility of measuring a given concept through some specified indicators is not hitherto guarantee of its objectivity nor of its adequacy.

With respect to this last remark, Reiss (2001) raises an interesting point. He discusses the economists' tendency of regarding as measurable any quantity for which there exists a measurement procedure. The problem, as Reiss points out, is that in case we have two different measuring procedures for the same concept, there is nothing in the economic theory that suggests why either procedure would be correct to use. He then distinguishes between natural and non-natural economic quantities. The former are those whose behaviour is described by our causal laws and which can be measured in a nonambiguous way, and the latter, by contrast, are either free-standing quantities (i.e. not embedded in the nexus of causal laws) or non ambiguously measurable. Clearly, the problem of isolating relevant concepts and of finding meaningful measurement procedures is not only a problem for economics but for all the social sciences and for the health sciences too (see, respectively, Zeller and Carmines 1980; Larson 1991).

The conceptual difficulty related to concepts and indicators has also practical pitfalls. For instance, as the health care costs rise, it is important to find correct measures of health to design effective health policies. However, the inadequacy of economic indicators was already denounced in the aforementioned paper by Land (1983), that gave a survey on the development of social indicators from the 1960s to the 1980s. Moreover, the dissatisfaction with the traditional quality-of-life constructs—i.e. with the *economic* measures (e.g., national income, gross domestic product, etc.) of quality of life—led social scientists and policy analysts to develop a new framework called *social quality* (Philips and Berman 2003). This framework has been developed since the late 90s for assessing the economic and social progress of groups, communities and societies. The motivation for such a development is in

the downgrading of social and cultural elements. The social quality approach, instead, bestows great importance on social justice, human dignity and participation.

Philips and Berman accept the definition of 'social quality' as "the extent to which citizens are able to participate in the social and economic life of their communities under conditions which enhance their well-being and individual potential", which has been proposed by Beck et al. (2001: 7). Thus social quality embraces the following elements:

(i) The degree of social security.
(ii) The level of social inclusion.
(iii) The extent of social cohesion.
(iv) The nature of national social cohesion.
(v) The extent of empowerment.

It is worth noting that this approach to social quality also has an important multilevel dimension insofar as it aims to assess the social quality of citizens in relation to their dealings with the nation or society. To measure this environmental dimension, Philips and Berman propose a series of indicators to use in individual and comparative studies. These indicators fall into four categories:

(i) Input indicators. They relate to resources available to the community, including legislation or the availability of services.
(ii) Process indicators, e.g. the eligibility and the extent of utilisation of services.
(iii) Outcome indicators. These address, for instance, the proportion of community needs that are met by providing these services.
(iv) Impact indictors. These relate to any resulting change the community.

What is more, the choice of concepts and indicators is closely related to the problem of *measurement*. The issue of measurement in the social sciences is addressed, for instance, in Zeller and Carmines (1980). The goal of their work is to convey the logic underlying measurement strategies and "more generally, to sensitize researchers to the multiple sources of non-random measurement error that probably influence data in the social sciences" (Zeller and Carmines 1980: Preface).

At the very beginning of the first chapter, Zeller and Carmines (1980: 1) claim that our capacity to answer research questions highly depends on the power and robustness of our measurement procedure. However, in bestowing so much importance to the measurement procedure isn't the theorising stage overlooked? In the same vein, Blalock emphasised the role of measurement considerations for theoretical purposes, in particular for considering new variables. The measurement stage thus seems to have priority over the theorising stage. But shouldn't it be the other way round? How are, or should, the theorising and measurement stages be intertwined? An obvious problem—not so obvious as to be omitted here, though—is that concepts have to be at a certain level of abstraction but not too abstract, otherwise there will be no correspondent variables that empirical researchers can employ. However, concepts provide a degree of generality and abstractness that theories do require if they are to be relevant in different historical periods and in different research settings. In other words, the choice of concepts and indicators is central also for *theorising*—that is generalising from findings of particular studies.

3.3.6 Causal Directionality

Let us assume, for the sake of the argument, that relevant concepts and indicators are correctly singled out. We now have to formulate the causal hypothesis. How do we choose causal directionality? How can we correctly specify the interrelations—i.e. the causal mechanism—between concepts? Time, in some cases, can be of some help as it also gives us the direction of the causal relation.

Once more, echoing Wunsch, God seems to have given the easier task to the natural scientists. Physical and physiological mechanisms, for instance, are themselves embedded in time: smoking at t causes lung cancer at t' ($t < t'$) but not the other way round. Causal directionality, however, is not so easily determined when *social* mechanisms are involved. For instance, does migration cause marriage dissolution, or the other way round? Where do we gather information to decide about causal directionality in this case? If we had theory—in this case, a demographic theory of migratory behaviour, or a sociological theory of marriage dissolution, or both—we could establish causal directionality on this basis. Unfortunately, as we have just seen, the social sciences can rarely take advantage of such systematised and general knowledge. If theory isn't of much help, perhaps time will be. We might hypothesise that migration causes marriage dissolution as the former is observed before the latter. This causal relation would make sense; however, it might be eventually disproved because, say, we overlooked a temporal prior process—e.g. marital problems and the subsequent decision to divorce—causing migration. Causal directionality is thus opened up to threats due to confounding and control, in turn due to problems of observability and/or lack of theory.

3.3.7 Searching Causes of Effects or Effects of Causes?

So far we have used the term 'causal relation' in quite a loose way. Causal directionality poses a problem because time ordering is not a guarantee, in many social mechanisms, of a correct causal ordering specification. But there is a second reason why the term 'causal relation' is somewhat vague: it does not specify, on its own, whether we search for causes of effects or for effects of causes. An established tradition, particularly related to the Neyman-Rubin-Holland approach presented in section 3.1.4, is devoted to the measurement of *effects of causes*, rather than explicating *causes of effects*. Similarly, Philip Dawid (2000, 2007) puts particular emphasis on the same aspect. In a nutshell, according to Dawid, inferences about causes of effects are beset by ambiguities of the counterfactual framework, whereas inferences about effects of causes do not require recourse to counterfactuals. Inferences about effects are *hypothetical* and thus testable, whilst inferences about causes are *counterfactual* and thus untestable. In fact, in a counterfactual conditional the antecedent is, by definition, contrary-to-facts, i.e. non observed. On this grounds Dawid strongly criticises any attempt to infer effects of causes exactly because they are based on nonobservable statements. Although this critique to counterfactuals has some plausibility, does it

follow that the endeavour to find out causes has to be dismissed altogether? Isn't the trouble with the counterfactual view that it *presupposes* which variables are the causes?

Another difficulty, as D. T. Cox (2000) correctly points out, is that the interpretation of retrospective studies is more hazardous than the corresponding interpretation of prospective studies. This is due, at least partly, to problems with the possibility of measurement. However, does this rule out as a sensible research programme to look for causes of effects? The opposite view is undertaken, for instance, by Heckman (2005), who clearly states that "science is all about constructing models of the causes of effects" (Heckman 2005: 2). And a few pages later (Heckman 2005: 6):

> The goal of the econometric literature, like the goal of all science, is to model phenomena at a deeper level, to understand the causes producing the effects so that we can use empirical versions of the models to forecast the effects of interventions never previously experienced, to calculate a variety of policy counterfactuals, and to use scientific theory to guide the choices of estimators and the interpretation of the evidence.

On the one hand, this seems to defend causal modelling from the hard attacks of those that, like Dawid, pretend to solve the problem of causal inference altogether by simply replacing it with decision theory. On the other hand, of course, Heckman's remark leads us back to the problem of theorising. But theorising is, on its own, a knotty issue in causal modelling. Sections 4.3 and 6.1 are a first step forward in better understanding the internal dynamic of causal models. In those sections I try to clarify the specific features that allow the causal interpretation—this goes towards a methodological improvement of causal modelling, rather than a drastic eradication of it from scientific practice—and I attempt a characterisation of causal modelling as the modelling of causal mechanism—again, this goes towards a methodological improvement of the goals of causal modelling.

3.3.8 Underdetermination

Another problem that causal modelling faces is that of *underdetermination*. In the philosophical jargon, the term indicates a particular relation between evidence and theory. Notably, many theories can account for the same evidence. As a consequence, theory cannot be warranted by evidence. Traditionally, underdetermination is debated in the context of natural science, for instance in physics or quantum mechanics. In the context of causal modelling, this means that several causal models can account equally well for the observed correlations. This is, needless to say, just a way of restating the uncontroversial claim that correlation is not causation and that, consequently, we cannot deduce causation from correlations alone—the problem was pointed out already by pioneers of quantitative social research such as Simon and Blalock.

But how is undertermination related to the H-D methodology discussed above? Psillos (2005) distinguishes two types of underdetermination: deductive and inductive. According to the deductive undertermination thesis, evidence cannot prove a theory to be *true*, whereas, according to the inductive underdermination thesis,

evidence cannot prove a theory to be *probable*. Deductive underdetermination rests on the claim, Psillos says, that the link between evidence and theory cannot be deductive, namely evidence cannot determine theory, which, in turn, seems to undermine the hypothetico-*deductive* character of causal modelling. However, I rejoin Psillos when he clarifies the claim that deductive undertermination besets an *unsophisticated* version of hypothetico-deductivism, rather than dismissing it altogether. Section 3.2 above should have already given an idea of the non-simplistic character of hypothetico-deductive causal models, and section 4.3 below will analyse in some detail the peculiar features of hypothetico-deductive causal models. However, the complexity of causal models does not just come from the complexity of their methodology but also from the complexity of their object of analysis itself: causal structures.

3.3.9 Complexity of Causal Structures

As we have seen in Chapter 2, the main limit of probabilistic theories of causality is to fall short when it comes to analyse more complex relations. An attempt to overcome this drawback of simple probabilistic accounts comes, I suggested, from causal modelling itself.

Early discussions of complex causal structures are already present in the pioneering works of Boudon (1967) and Boudon and Lazarfeld (1966) in the scientific literature, and of Mackie (1974) in the philosophical literature. Franck (2003) offers a bright account of the methodological and epistemological gains in abandoning the rigid framework of 'same cause, same effect' and provides a summary of alternative structures. In a nutshell, causal analysis has to take into account two possible patterns: (i) A plurality of causes brings about the effect; and (ii) a cause brings about different effects. Franck presents the following scenarios:

(i) Disjunctive multiple causation: several causes can produce a single effect, each cause operating independently of each other.
(ii) Conjunctive multiple causation: two or more causes together produce an effect, though each alone does not.
(iii) INUS conditions: a cause is defined as the Insufficient but Necessary part of a condition which is itself Unnecessary but Sufficient for the effect.
(iv) Interactions: a single cause produces different effects, being associated with other causes.

However, this list omits other scenarios:

(v) Causal chains: a cause C produces an effect E which in turn is a cause of a different effect F.
(vi) Causal loops: a cause C produces an effect E that in turn influences C but at a later time.
(vii) Dynamic causal structures: the analysis of causal structures that change over time (the causal mechanism changes altogether).

It goes without saying that finding successful methodologies to tackle complex causal structures is important both for cognitive reasons (having a good understanding of the causal mechanisms) and for action-oriented reasons (detecting which factors to intervene upon). However, to have a good inventory of various causal structures at our disposal is not enough. The problem of choosing the correct causal structure is exactly the problem of causal modelling—that is the modelling of the causal mechanism. This problem will be addressed in Chapter 6.

3.3.10 Interpretation of Tests

Again for the sake of argument, let us suppose that we come up with a plausible causal hypothesis. The testing stage is unfortunately no less problematic than the hypothesis formulation stage. Simply put, the goal of hypothesis testing is to compare the hypothesis with observations sampled from the population. A statistical test involves the following elements. The *null hypothesis* states that an observed difference just reflects chance variation. The *alternative hypothesis* states that the observed difference is, instead, real. The rejection of the null hypothesis depends on the *test statistic* and on the *significance level*. The significance level is chosen on the basis of the amount of type I or type II error one is prepared to accept and on the basis of the problem at hand.[4] The test statistics measures the difference between the data and what is expected under the null hypothesis. Typically, three test statistics are used: the z-test, the F-test, and the X^2-test. The null hypothesis is said to be accepted or rejected at a given *significance level*—the *p-value*—usually set at 5%. This, in a nutshell, is a test. For a very lucid and accessible presentation, see Freedman et al. (1998: chaps. 26–29).

Although intuitively simple, tests of significance hide a number of interpretational difficulties. Freedman et al. (1998: ch. 29) and Freedman (2005: 60 ff.) provide us with a good inventory of the tricky issues. To begin with, the word 'significance' might be misleading as in the statistical jargon it is not synonymous with 'important', 'relevant', but with 'probably true', i.e. not due to chance. Secondly, the p-value of a test depends on the sample size and presupposes the quite strong assumption that the sample is representative of the population. Also, the threshold for significance is rather relative. Standard practice prescribes the rejection of the null hypothesis at the 5% or at 1% level. However, those levels are arbitrary. Freedman et al. (1998: ch. 29) make the point that there isn't a real difference between two p-values, say one set at 5.1% and the other at 4.9%. Therefore they recommend reporting the test used and the exact p-value, or otherwise 'statistically significant' will be too much of a vague statement. But the p-value raises other interpretational problems. Lagiou et al. (2005), for instance, discuss the interpretation of p-values in the context of epidemiological research and points to two major difficulties:

[4] A type I error is made if the null hypothesis is rejected when in fact it is true, and a type II error is made when the null hypothesis is accepted when instead the alternative is true. On the two types of errors and their relations with the interpretation of probability, see also section 5.4.

(i) the *p*-value is interpretable only when one comparison or one test is performed; and (ii) the *p*-value itself does not convey information about the strength of the respective association. The *p*-value does not give the chance of the null hypothesis being true. A small *p*-value has to be interpreted as evidence against the null hypothesis, in particular as suggesting that something beside chance is operating to make the difference. More to the point, as Freedman et al. (1998: ch. 29) again explain brightly, a test of significance cannot shed light on the causes of variations—it just tests whether an observed variation is real (alternative hypothesis) or just chancy.

It is worth noting that hypothesis testing is in fact about testing *variations*. This interpretational aspect is often neglected in the literature, and in section 4.1 we will see, instead, how fundamental the notion of variation is for causality. Nonetheless, even though statistical procedures allow us to get a grip on significant variations, there will always be a margin of uncertainty due to measurement errors, missing data, sample size, etc.—the null hypothesis is usually rejected at the 5% level, but this does not rule out the possibility that the variation was instead chancy. This probability is tiny, but still exists. For this reason practising scientists often decide which hypotheses will be tested after having seen the data. Freedman (2005: 64) calls this practice *data snooping*. Data snooping is used, for instance, to exclude insignificant variables. But this is, as we have seen earlier, an important step in the hypothetico-deductive methodology of causal models. The *causal* interpretation then goes far beyond the rejection of the null hypothesis, and section 4.3 will thoroughly discuss the peculiar features of causal models that allow the causal interpretation of statistical procedures.

Another problem in hypothesis testing concerns the interpretation of probability. On the one hand, from a frequentist point of view, we are not concerned with the probability of a hypothesis being true, but with the probability of obtaining the observed sample *if* the hypothesis is true. The frequentist is in fact unable to attach a probability value to a single case—such as a hypothesis—because probability expresses frequency of occurrence in finite or infinite sequences. On the other hand, Bayesians can attach a probability value to the single case and therefore have a meaningful way of expressing the probability of a particular hypothesis. I direct the reader to Courgeau (2004b) for a very clear account of the meaning of hypothesis testing in a frequentist and in a Bayesian framework. The Bayesian framework, and particularly objective Bayesianism, will be the object of Chapter 5, where I shall address the problem of the adoption of the interpretation of probability; in the concluding remarks of the book I shall discuss the consequences of such an adoption on methodology.

3.3.11 Validity

Finally, we come to the validity of results. When is a causal hypothesis confirmed? Section 4.3 will be devoted to the vindication of the causal interpretation of

statistical models. However, the approach taken in section 4.3 needs to be completed with an analysis of the meaning of validity in the context of statistical modelling. A clear account is that of Cook and Campbell (1979). The merit of Cook and Campbell is to discuss thoroughly the possible threats and difficulties that undermine validity. They borrow the terms 'internal validity' and 'external validity' from an earlier work by Campbell and Stanley (1963), and add two subcategories, namely statistical validity and construct validity.

To begin with, the terms 'validity' and 'invalidity' refer to the best available approximation to the truth of causal statements. Cook and Campbell list four fundamental questions that need to be answered in any applied research:

1. Is there any relationships between two variables?
2. Given that there is one, is the relationship plausibly causal from one to another, or would the same relationship have been obtained in the absence of any treatment of any kind?
3. Given that the relationship is plausibly causal, what are the particular cause and effect constructs involved in the relationship?
4. Given that there is a relationship from construct A to B, how generalisable is the relationship?

The four questions are, respectively, questions about statistical, internal, construct, and external validity.

Statistical validity ascertains whether the study is sensitive enough to permit reasonable statements about covariation; statements about covariation are necessary for any statement about causality—I shall get back to this point shortly. Once statistical validity establishes with reasonable confidence a co-variational statement, *internal validity* establishes whether the relation is causal, or whether from the absence of a relationship between the two variables we can infer absence of causality. As has been discussed earlier, a difficulty of causal modelling is to choose appropriate concepts and indicators, and, accordingly, to use correct constructs to measure them. *Construct validity* is then concerned with the approximate validity of statements that involve such constructs. For instance, a particular cause or effect can be represented in terms of different constructs and the goal will be to ascertain the validity of alternative statements. Finally, *external validity* concerns the possibility of generalising a presumed causal relationship across different times, settings, populations. In other words, given a particular study, this will be statistically valid if statements about covariation can be made with reasonable confidence, internally valid if a causal relation is confirmed within the specific population at hand, constructively valid if alternative constructs for causes and effect deliver consistent results, and it will be externally valid if the results can be generalised to other populations.

Threats to statistical, internal, construct and external validity are of a different nature. Statistical validity, for instance, can be threatened by the sample size or the heterogeneity of individuals in the population. In assessing internal validity, time plays an important role. In non-experimental contexts knowledge of time ordering

of variables might and arguably should participate in the determination of the direction of a causal relation. However, this does not rule out the possibility of mistakenly inferring $A \rightarrow B$ when in fact there is an intermediate variable, the correct path being $A \rightarrow C \rightarrow B$. The gravity of such an error depends, of course, on the role C plays in the mechanism. Also, internal validity depends a great deal on the sources of random error and on the appropriate use of statistics and of statistical tests. Construct validity has the obvious problem of the underrepresentation of constructs, and external validity largely depends on the representativeness of samples and on the possibility of multiple replication of studies. Cook and Campbell provide a clear and informed discussion of various issues, especially related to randomization, representativeness of samples and testing. I will not linger extensively on all threats here—some threats have already been discussed earlier in this section on the difficulties and weaknesses of causal modelling, and others will be presented in section 4.3.

In the following, I wish to draw the reader's attention to three issues. The first concerns the ultimate content of validity statements. I briefly mentioned above that statistical validity assesses whether covariation between two variables holds and that such a statement is necessary for causality statements. Let me now elaborate more on that. This claim—that "Covariation is a necessary condition for inferring a cause" (Cook and Campbell 1979: 37)—might look trivial on its own, but it is not. Recall, we are here looking for the *rationale* of causality in causal modelling and this claim puts us on the right track. *Covariation* about variables—not regularity, nor invariance, but joint *variations*. Section 4.1 will explain the fundamental role the notion of variation plays in causal modelling and will contrast it with competing views that see, instead, in regularity or invariance the ultimate test for causal inference. In particular, it will be shown that variation conceptually precedes invariance and regularity which are instead constraints on the joint variation.

The second issue concerns the priority of validity. Cook and Campbell (1979: 82–85) suggest that the order of interest for theoretical research probably is internal, construct, statistical and external validity. However, the goal of applied research being different, the order of priorities in this field will also be different, viz. internal, external, construct validity of the effect, statistical and construct validity of the cause. The choice of the orders is arguably not an objective one and researchers with different backgrounds, interests or goals might suggest different priority lists. Also, the choice of the priority list might depend on the researcher's view of other issues—for instance of what theorising is or of the possibility of establishing social laws.

Finally, although Cook and Campbell's presentation and discussion of types of validity is extremely elucidative about the levels at which we can test a causal model—that is, its statistical power, its internal validity, the reliability of its construct, or the possibility of generalising causal relationships— they stay silent about the ultimate goal of causal modelling. In section 6.1, I will advance the view that the ultimate goal of causal modelling is to model causal mechanisms where certain variational relations among variables of interest hold.

3.3.12 Generalisability

Finally, we face the problem of generalising results. As Cook and Campbell (1979) correctly noticed, we should distinguish between extending results from the sample to the population, which is a matter of internal validity, and extending results to different populations, which is instead a matter of external validity. According to their account, the main threat to generalisability is the representativeness of samples. This is exactly true but, I'd add, on what *theoretical* grounds can we extend results?

The problem of generalising results is related, for instance, to the problem of finding or choosing appropriate or suitable concepts. These concepts, as we have seen, have to be abstract enough to permit a generalisation over space and time. However, this is no guarantee at all. For instance, in discussing several studies in the field of child mortality in developing countries, Masuy-Stroobant (2002) points exactly to this difficulty. In particular, Masuy-Stroobant claims that unlike the physical and medical sciences that use standardised concepts and indicators, the social sciences have to deal with a great heterogeneity and with a lack of comparability both over studies and over time. The problem is probably even more serious if we consider qualitative research but this, according to Masuy-Stroobant, can be embanked by organising qualitative observations in a quantitative framework.

In other words, the problem of generalisability of results highly depends on the possibility of finding concepts and constructs that are applicable and comparable across different studies in different populations. This, in turn, is a problem of level of abstraction. However, we can determine the level of abstraction of concepts and constructs only if we provide the statistical model with some non-mathematised contextual and background knowledge. Section 4.3 will explain why background knowledge is so important in causal modelling and will attempt a more precise characterisation of this notion.

We now have at our disposal all the necessary methodological tools for building an articulated epistemology of causal modelling. This is the task of next chapter.

Chapter 4
Epistemology of Causal Modelling

Abstract This chapter develops the rationale of causality. It is argued that causal models are regimented by a rationale of variation, not of regularity or invariance. Namely, causal models establish causal claims by evaluating suitable variations among variables of interest. It is also argued that regularity and invariance are constraints to impose on variations in order to guarantee their causal interpretation. Empirical, methodological, and philosophical arguments are offered. A taxonomy of variations is also sketched, and a thorough comparison between associational and causal models is offered in order to pinpoint the features of causal models guaranteeing the causal interpretation.

Keywords Rationale of causality; variation; taxonomy of variations; causal interpretation; assumptions; background knowledge; invariance and stability; associational models; causal models; foundations of causal analysis.

Introduction

We now come to the central part of the book and thus to its central thesis. Chapter 3 closely investigated the methodology of causal modelling and prepared the ground to develop its epistemology. Epistemological issues in causal modelling concern, on the one hand, *how* we come to know about causal relations, and, on the other hand, *what allows* us to know that a relation is causal. More specifically, we will be here concerned, respectively, with the rationale of causality and with the vindication of the causal interpretation of models.

Section 4.1 develops the rationale of causality. The idea is that causal models are regimented by a rationale of variation, that is they are tested by the evaluation of relevant *variations*. Therefore, the notion of *variation* is discussed at length, and it will be explained how this rationale underlies causal models presented in Chapter 3. I argue that the rationale of variation breaks down the received view that sees in regularity and invariance the bottom-line concepts for causality. I support the

F. Russo, *Causality and Casual Modelling in the Social Sciences: Measuring Variations* 91
Methodological Prospects in the Social Sciences

rationale of variation by offering methodological as well as philosophical arguments and by showing (i) that the rationale was already employed by the forefathers of quantitative causal analysis, and (ii) that it is also latent in the contemporary philosophical literature. Finally, I provide further evidence by discussing possible objections. Failing to recognise the rationale of variation—I claim—leads to dangerous conceptual confusions.

Section 4.2 outlines a taxonomy of variations according to five criteria: variations (i) across time, (ii) across individuals, (iii) across characteristics, (iv) counterfactual and control group variations, (v) observational vs. interventional variations. In fact, the presentation of the variation rationale in section 4.1 is still too general, and a scrutiny of case studies shows that social scientists look for different types of variations, depending on the case at hand.

Section 4.3 confronts us with the problem of the warranty of the causal interpretation, i.e. what allows us to *interpret* variations causally. I will carry out a careful examination of features of associational models and of causal models. It will be shown that the latter have supplementary sets of assumptions and the fundamental role of those additional assumptions for the causal interpretation will be explained.

4.1 The Rationale of Causality: Measuring Variations

The presentation of a few case studies in Chapter 1 gave us the flavour of what causal analysis in social science looks like; and after this presentation, I drew some preliminary methodological and epistemological morals. In particular, I emphasised (i) that social scientists opt for statistical causality rather than aleatory causality, (ii) that a causal claim is stated in a causal context and then put forward for empirical testing, (iii) that social scientists look for specific variations to be tested, and (iv) that causal mechanisms are modelled by means of statistical tools. Then, we asked philosophy to give us an account of causality that would fit the statistical characterisation of causality in social science, and we discussed the limits of those philosophical attempts to provide a meaningful scheme of scientific practice. We saw that although Good's and Suppes' approaches proved not to be powerful enough to isolate the rationale of causality we are looking for, they allowed us to highlight an essential element of a statistical characterisation of causal relations: statistical relevance. The main drawback in those theories was to neglect the multivariate aspect of causality, which is taken into account, on the contrary, in causal modelling. Chapter 3 investigated at length the kind of statistical notions used in quantitative causal analysis. In particular, the multivariate aspect of causality is mirrored in the fact that causal relations are expressed in equations—for equations are more informative than simple probability inequalities—and also in the fact that, to describe complex social phenomena, we often use *systems* of equations. The analysis of the methodology of causal modelling prepared the ground for a thorough examination of other epistemological issues, in particular, social scientists look for meaningful *variations* to test among variables of interest.

The task is now to carry on—and to go beyond—the work initiated with probabilistic theories of causality and with causal models in the previous two chapters. More specifically, an epistemological account of causality in causal modelling aims at providing a deep understanding of which *notion* of causality social scientists currently employ in causal analyses. Such an epistemological notion of causality represents, in other words, the *scheme* of causal reasoning, i.e. it is a *rationale*. It is worth noting that to provide a rationale does not mean to give a definition of what causality is.

A *rationale* is the principle or notion underlying some opinion, action, hypothesis, phenomenon, reasoning, model, or the like. The quest for a rationale of causality in causal modelling is a search for the concept that guides causal reasoning and thanks to which we can draw causal conclusions. The search for the rationale of causality underlying causal models is the investigation of the concept of causality employed in those models, whether explicitly or implicitly. Consequently, to find a rationale of causality is, in the first place, a job for epistemology with important consequences on methodology. In other words, to give a rationale means to provide the bottom-line concept for causality. To provide a definition of causality is, undoubtedly, a job for metaphysics, whereas we are here concerned with epistemology and methodology, and our privileged question concerns *how* we come to know about causal relations, not what causality in fact is. So, what kind of rationale regiments causal modelling? Is it a rationale of regularity and invariance?

There is a widespread belief that regularity and invariance provide the ultimate concepts of causality or the strongest statistical conditions to ensure the causal interpretation or model. On the one hand, philosophical theories still tacitly assume that causality is a matter of regularity, a heritage of Hume. On the other, the invariance condition is considered *the* condition allowing the causal interpretation, a heritage of developments in the econometric literature. In the following, I will challenge both claims—that causality is in regularity and that causality is in invariance—and I will advance the view that the rationale of causality in causal modelling is, instead, a rationale of variation.

Let me open the discussion by quoting Hubert Blalock, a pioneer of quantitative causal analysis. Blalock (1964: 9–11) discusses the concept of causality related to the concept of production:

> If X is a cause of Y, we have in mind that a change in X produces a change in Y and not merely that a change in X is followed by or associated with a change in Y.

This emphasis on the notion of production has been further echoed by Bunge (1979a). Yet Blalock acknowledges that an empiricist scientist would easily object that we cannot really observe or measure causes that *produce* effects. The difficulty is that we cannot test whether the change in X actually *produces* the change in Y, unless we postulate a causal mechanism responsible for that beforehand. *Producing* refers to an ontological process, whereas here we are interested in the epistemological principle—i.e. in the rationale—that allows causal inferences in causal modelling. This principle cannot be in production: the ontological process—the causal mechanism—is assumed (see later section 4.3) or modelled (see later

section 6.1) and therefore it cannot be what guides causal reasoning in causal modelling. Consequently, whether or not causality involves *production* is a metaphysical issue that I shall leave aside. Nonetheless, Blalock puts us on the right path, because, although he stresses the fact that causes *produce* effects, he also says that a change or *variation* in the causal variable produces a change or *variation* in the effect variable. It is my purpose to investigate further this concept of change or variation.

The study of change is thus the study of factors which produce change. This claim, however, still sticks to the notion of production rather than variation. Moreover, from an epistemological viewpoint, we ought to consider carefully that our epistemic access to the data generating process *de facto* passes through correlational data. Of course, the best research technique would be experimentation, for experimentation allows manipulation and control, hence causal variations would be more easily recognizable. Experimentation is typically advocated as the preferred tool for causal analysis for two reasons: (i) units of interest can be experimentally manipulated (for instance by random allocation); and (ii) experiments, being in principle repeatable, allow generalisation of results over space and time. Nevertheless, social studies often have at their disposal only observational, but not experimental, data. Notably, the difficulty in observational studies is how scientists can extract the best description and explanation of variations from mere observations and, from that description, how they can draw causal conclusions. The answer seems to lie in a careful scrutiny of the notion of variation.

Contra the received view, I advance that this epistemological principle—the rationale of causality in causal modelling—lies in the notion of *variation*. More specifically, when we say that X is a cause of Y, we mean that a *variation* in Y has been detected, and that we have good reasons to say that X, particularly, a change in X, is responsible for, i.e. produced in Blalock's words, the change in Y. Epistemology and methodology investigate what these 'good reasons' are, and they are not satisfied with just postulating a metaphysical causal mechanism.

Let us go one step further. Measuring variations conveys the following idea: to test causal models means to measure suitable changes or variations. As I pointed out in section 3.2, causal models apply a hypothetico-deductive methodology: causal hypotheses are first formulated, and then put forward for empirical testing. In turn, empirical tests are designed to assess the presence of a variation, and to assess whether this variation satisfies certain conditions, for instance invariance. Moreover, standard hypothesis testing tests whether a *variation* is chancy (null hypothesis) or real (alternative hypothesis). Different types of variations and their characteristics will be discussed in section 4.2 in this chapter. Let me now spell out in detail the rationale I propose.

4.1.1 Methodological Arguments

To explain at length the variation rationale, I will go through probabilistic theories of causality first, then through structural equation models, covariance structure models,

contingency tables and, at last, through the double-frontier approach. The objective is to show how a rationale of variation is in fact involved in them.

Probabilistic theories of causality. Probabilistic theories of causality focus on the difference between conditional probability $P(E|C)$ and marginal probability $P(E)$. To compare conditional and marginal probability means to analyse a statistical relevance relation. The underlying idea is that if C is a cause of E, then C is also statistically relevant for E. Hence, the change hereby produced by C in the effect E will be detected because the conditional and the marginal probability *differ*. Analogously, quantitative probabilistic theories focus on the difference between the conditional distribution $P(X \leq x|Y \leq y)$ and the marginal distribution $P(X \leq x)$. Again, to compare conditional distribution with marginal distribution means to measure the *variation* produced by the putative cause X on the putative effect Y. In the ultimate analysis, probabilistic theories of causality, by using statistical relevance as a necessary ingredient, establish causal relations by making *variational* claims, that is the *variation* in the conditional probability of the effect is due to a *variation* in the marginal probability of the cause. It is worth noting that whilst variation is a necessary condition to infer causality regularity isn't. Probabilistic theories of causality might require regularity to allow generalisations, but not to infer causality *tout court*. More to the point, the evaluation of single-case relations can be done by comparing the marginal and the conditional probability of the event, without requiring a regular instantiation of the causal relation.

However, to claim that we expect a cause to affect the marginal probability of the effect is not equivalent to saying that from independence we can straightforwardly conclude to the absence of a causal path, for there are obvious cases in which this is not true. In Bayesian nets terminology, the model might be unfaithful, although causal. In the field of Bayesian networks, the faithful condition states that there are no probabilistic independencies apart from those that are consequences of the Causal Markov Condition. This phenomenon is apparent in Hesslow's example about thrombosis, pregnancy, and contraceptive pills (Hesslow 1976) (see Fig. 4.1).

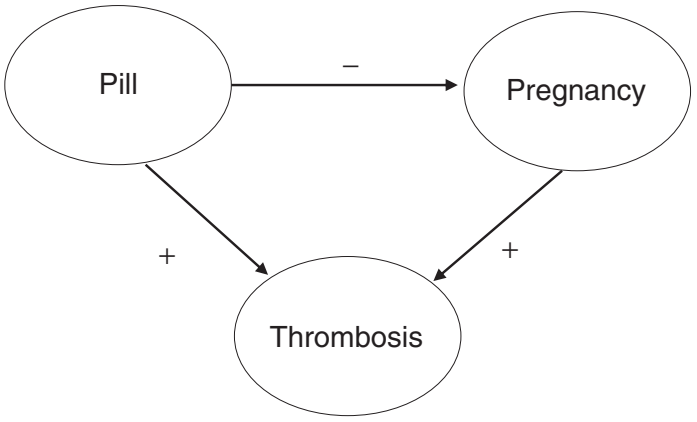

Fig. 4.1 Graph for Hesslow's example

If the two paths $pill \xrightarrow{+} thrombosis$, $pill \xrightarrow{-} pregnancy \xrightarrow{+} thrombosis$ exactly balance then there is no probabilistic dependence between *pill* and *thrombosis*. Yet, according to the graph, there is a path, so the model is unfaithful. This simple example also shows the importance of distinguishing between direct and indirect paths from causes to effects.

Structural equation models. Consider now structural equation models (SEM). Recall that the basic idea of SEM is that in a system of equations we can test whether variables are interrelated through a set of linear relationships, by examining the variances and covariances of variables. Sewall Wright, as I mentioned earlier in section 3.1, has taught us to write the covariance of any pair of observed variables in terms of path coefficients and covariances. The path coefficient quantifies the (direct) causal effect of X on Y; given the numerical value of the path coefficient β, the equation claims that a unit increase in X would result in β units increase of Y. In other words, β quantifies the *variation* on Y accompanied by the *variation* on X.

Another way to put it is that SEM attempt to quantify the change in X that accompanies a unit change in Y. As we have already noticed, the equality sign in structural equations does not state an algebraic equivalence; jointly with the associated graph, the structural equation is meant to describe the causal mechanism implied by the data generating process. The path coefficient β, in turn, is meant to quantify the (direct) causal effect of X on Y. This is why β quantifies the *variation* on Y produced by the variation of X.

Here is another reason why the notion of variation really matters for causation. It is commonly agreed that in structural equations variations in the independent variables, i.e. the Xs, (statistically) explain the variation in the dependent variable, i.e. Y. Assuming that covariate sufficiency holds, all and only the Xs included in the model explain the variation in Y. This is not a new idea—as Haavelmo (1944: 23) put it as early as 1944:

> […] for each variable, y, to be "explained," there is a relatively small number of explaining factors the *variations* of which are practically decisive in determining the *variations* of y. (My emphasis.)

The interpretation of the structural equation involves the notion of variation, not of regularity. In fact, the equation states that variations in X are accompanied by variations in Y, not that values of Y regularly follow values of X. Moreover, invariance, as we shall see later, is a constraint on the variation, not a condition of regular succession.

One might suggest that the explanatory power of a causal model is given by the inverse of the unexplained variance in the dependent variable. Namely, because the dependent variable is determined to a certain extent by variations in the independent variables, the more we can account for those variations, the higher the explanatory power of the model. What is not determined by variations in the independent variables depends on the errors. In fact, the coefficient of determination r^2 measures the portion of the variance in the dependent variable the independent variable is responsible for, or, put differently, the squared path coefficient represents the proportion of common variation in both the dependent and independent variables. The squared

path coefficient is unity only in case all assumptions of independence hold, in which case the variance in Y is completely explained by the variances of the Xs. This is seldom the case, however. There is another methodological threat, though. r^2 measures the goodness of fit, not the validity of the any underlying causal model. This means that the guarantee that we hit upon a correct causal structure has to be assessed on other grounds. We shall examine these other grounds later in section 4.3 in this chapter. The theoretical insufficiency of high r^2 values will also be discussed in section 6.1.

Covariance structure models. The rationale of causality as the measure of change emerges even more clearly once we examine covariance structure models (CSM). By way of reminder, CSM have two models: a measurement model and a structural model. CSM attempt to explain the relationships among a set of observed variables (in the measurement model) in terms of a generally smaller number of unobserved variables (in the structural model). Model building turns around this idea: covariances in the measurement model are explained by the (causal) relations as indicated in the structural model. To analyse covariances means exactly to measure (joint) changes, that is, to measure (joint) *variations*. This is definitively consistent with the basic idea of probabilistic causality, but also with structural equation models. Indeed, in CSM both ideas are combined.

Contingency tables. Causal analysis by contingency tables also relies on the rationale of variation. The key question is, in fact, the extent to which each of the independent variables contributes to the *variation* across the categories in the dependent variable. In the scheme of reasoning behind contingency tables, cunning philosophers will have already recognized the more familiar *statistical-relevance (S-R) model* of explanation advanced by Wesley Salmon (1971, 1984). And in fact, not surprisingly, the rationale of variation is clearly involved in the S-R model as well. According to Salmon, to explain a fact, i.e. to identify its causes, one has to identify the correct cell in the reference class the fact to be explained belongs to.[1]

Consider Salmon's example (Salmon 1984: 37). To understand why a particular individual—say, Albert—committed a delinquency—say, stealing a car—we first consider the broadest reference class Albert belongs to (American teenager), we then partition this class into subclasses corresponding to the number of (all and only) relevant factors. Sociological theories suggest taking gender, religious background, marital status of parents, type of residential community, socio-economic status and education, etc. into account. We will thus obtain a large number of cells, each of which will be assigned a probability of the degree of delinquent behaviour. This fact, i.e. why Albert committed a delinquency, will be explained once the *narrowest* class Albert belongs to is identified, e.g. male *and* parents divorced *and* living in a suburban area *and* low education.

[1] It is worth noting that later on Salmon (1984) changed his mind and argued that statistical relevance is not enough to infer causal relations. What is needed, instead, is a causal process from the cause to the effect. Salmon developed the statistical relevance model to offer a solution to the counterexamples raised against the high-probability requirement in the inductive-statistical model developed by Hempel.

As the name suggests, statistical *relevance* relations are used in the S-R model in order to isolate *relevant* causal factors. Let me explain how it works—the rationale of variation will then become apparent. Let A denote American teenagers, and B_i various degrees of juvenile delinquency. What we are interested in is not only $P(B_i|A)$, that is, the probability of committing a delinquency in the population of American teenagers. In particular, we are instead interested in a more specific probability, say $P(B_i|A \cap C_j \cap D_k \cap E_n)$, where C_j, D_k and E_n are all *relevant factors*, for instance gender, religious background, marital status of parents, etc. The crucial point is that if conditioning on a further factor, say F_m, does not *change* the previous conditional probability, this means that F_m is not a relevant factor and hence should not be considered in the explanation. So all factors entering the S-R model are statistically relevant, i.e. *responsible for variations*, in the probability of the fact to be explained.

Double-frontier approach. The rationale of variation is also (implicitly) used in an exploratory methodology called double-frontier analysis. The application of this methodology is the object of a current research project run at the Institute of Demography at the University of Louvain and is applied to evaluate the impact of the characteristics of the newborn and his/her health status at birth (Vandresse 2005, 2008). The double-frontier approach is borrowed from the field of econometrics, where it is widely used for the measurement of productive efficiency. However, it is completely new in quantitative demography. I shall first present very briefly the idea behind this approach, then explain why it is innovative for causal analysis in demography, and last show how the variation rationale is involved in it.

In the econometric literature, the efficiency of a firm is usually defined in terms of two distinct concepts: technical efficiency and allocative efficiency.[2] Technical efficiency is the ability of a firm to produce a maximum level of output from a given set of inputs; allocative efficiency is the extent to which firms use the inputs in optimal proportions. The relevant efficiency measure is then defined relative to a notion of best practice at a particular point in time. Measuring the efficiency will then require the computation or estimation of an efficiency frontier. A particular firm will be 'efficient' if it operates on the efficiency frontier, and 'inefficient' if it operates above the same frontier (see Fig. 4.2).

The classical approach to child mortality and morbidity (Mosley and Chen 1984) uses a structural framework where the statistical model is based on the expectation of the variable of interest conditional on a set of explanatory variables. The interpretation of results is thus based on the mean and on the deviation from this mean. In other words what the classical approach tries to detect is an *average causal relation* (see Fig. 4.3).

Clearly, in the classical approach extreme values are not considered. The double-frontier approach, instead, is mainly interested in those extreme values. Suppose we consider the weight at birth as a proxy of the health indicator. The question would then be: given certain characteristics, what is the maximum and the minimal weight the newborn can attain? Differently put, observations are evaluated with respect to

[2] For details on the frontier approach in econometrics see Fried, Schmidt, and Lovell (1993).

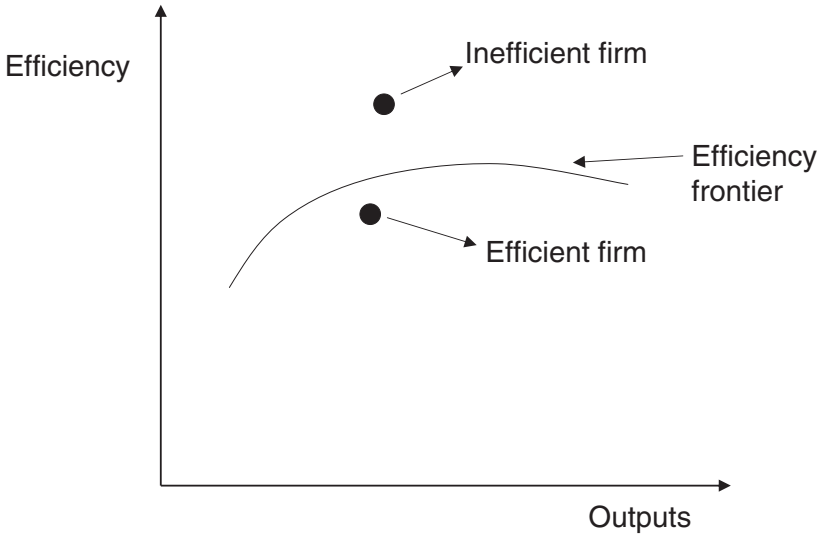

Fig. 4.2 The frontier approach in econometrics

extreme values rather than with respect to the mean value. This means computing or estimating two frontiers: a maximum health indicator frontier and a minimum health indicator frontier. Frontiers are indeed sensitive to the presence of outliers: they might influence the estimation of the frontiers as well as the distance between them. The problem of how to deal with outliers concerns the double-frontier approach as

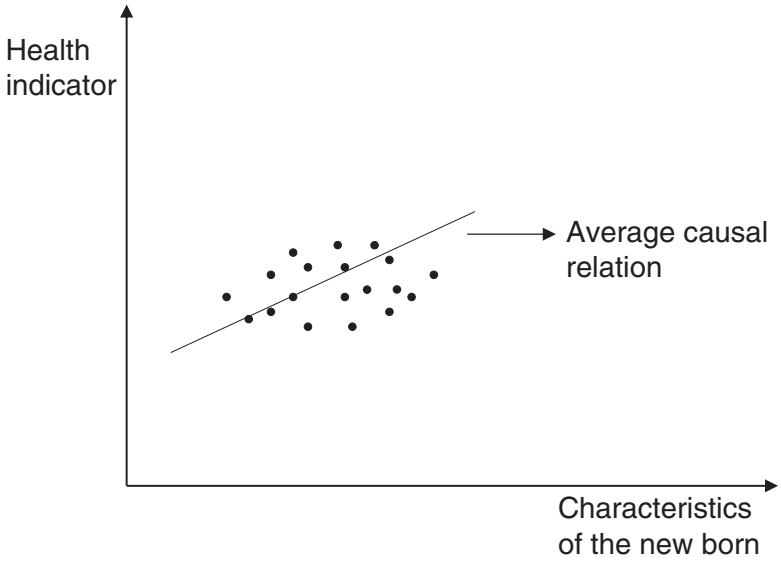

Fig. 4.3 The classical approach to child morbidity

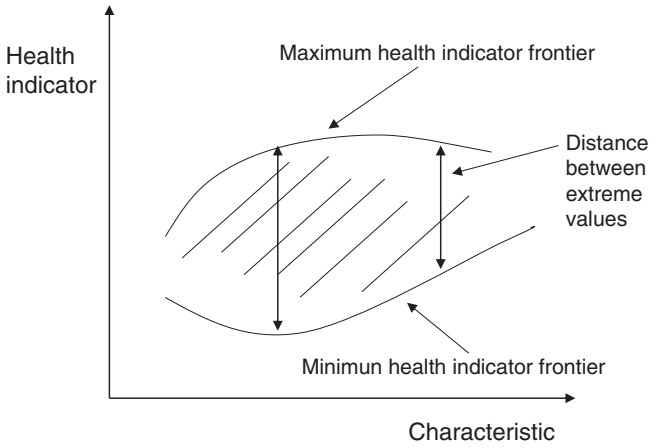

Fig. 4.4 The double frontier approach to child morbidity

well as traditional causal models. Most of the time outliers are excluded from the data set, but this might of course be a mistake.

In this approach, the causal impact of a certain characteristic is evaluated, roughly, depending on the concentration of the set of observations between the two frontiers. Simply put, the goal is not to estimate an average causal effect; instead, we are interested in the distance between the maximum frontier and the minimum frontier—i.e. how the health indicator *varies* between extreme values, and on how this distance *varies* for different levels of the explanatory variables. The shorter the distance between extreme values, the stronger the (causal) influence of the characteristic under analysis on the health indicator (see Fig. 4.4). For instance, for each level of maternal age the distances between upper and lower observations are relatively high; the distance increases up to 30 years of age and then decreases. The causal interpretation of such variations of the distances is that maternal age is not significantly relevant for the health indicator at birth up to 30 years of age, but it is after.

This approach is a turning point both from a methodological and epistemological perspective. Although even Quetelet, as we shall see later, employed the rationale of variation for discovering the 'laws of social physics' regimenting the 'average man', the double-frontier approach pays attention to *variations* in the first place, not to mean values, as is the case in standard causal modelling. This means that the rationale of variation conceptually precedes the condition of stability of the putative average causal effect—I shall argue more widely in favour of this conceptual precedence later in this section. The double-frontier approach reveals very clearly that what matters for causation in the first place is the evaluation of *variation*. This is not to say that standard structural modelling has to be thrown away; rather, I want to stress that (i) standard causal modelling and the double-frontier approach are complementary, and (ii) the variation rationale sheds light on the interpretation of both standard causal modelling and the double-frontier approach. A structural framework

is indeed necessary for the choice of variables to be analysed in the double-frontier model, and, at the same time, the double-frontier model allows us to get a more complete causal picture. Average causal relations tell only part of the causal story. In particular, average causal relations do not tell us anything about the upper and lower boundaries of the causal relation.

4.1.2 Foundations

Let us now go back in time. I will show that the rationale of variation is already present in the works of Adolph Quetelet, John Stuart Mill, Emile Durkheim, and also in some fairly recent foundational texts in epidemiology.

Adolph Quetelet. In the *Physique sociale* (1869: ch. 10, Book I, Tome I), Quetelet states very clearly the objective of his work: to study the causes—whether natural or disturbing—that operate on the development of man. In particular, the goal is to measure their influence and their mode of reciprocal action. Quetelet then spends a few paragraphs on explaining, quite informally, how he intends to detect and measure the causes that influence his *average man*. What he describes is the Millian comparative method. For instance, suppose we wonder what is a man's disturbing influence that modifies his physical force. Once we have obtained several scales of the force of the hands or kidneys of a great number of men, we will compare those scales, and particularly their mean values, in order to test whether the quantity of force decreased or increased by the disturbing action of the cause. This *variation*—says Quetelet—is what matters for a social physics.

Quetelet's search for causes of the human development started with the observation of a wide *variability* in the mortality tables he calculated, not of a wide regularity (Quetelet 1869: ch. 9, Book I, Tome I). The causes of the development of the average man are to be detected by a careful examination of the *variations* displayed in the mortality tables. Thus, his causal reasoning is pervaded, from the very start, by the variation rationale. It is worth noting that the entire enterprise in the *Physique sociale* is not merely descriptive—the causalist perspective is present and explicit from the very beginning.

John Stuart Mill. Mill, in the *System of logic* (1843: ch. 4, Book III), mainly talks about the cause as a *physical* cause. The cause, in this sense, is the whole range of antecedent conditions, positive and negative. Of course, this notion of cause underlies a strong experimentalist view. In the *System of logic* the experimental inquiry is seen as the solution to the problem of finding the best process for ascertaining what phenomena are related to each other as causes and effects. Causal analysis, says Mill, requires following the Baconian rule of *varying* the circumstances, and for this purpose we may have recourse to observation and experiment.[3] In this

[3] Mill (1843: Book III, chapter VII, §2) specifies that between observation and experiment there is no logical distinction but only practical. The problem of causality in observational studies is not overtly addressed.

general idea of *varying the circumstances* in order to establish cause–effect relations, the variation rationale is clearly already at work. Let us now look at the comparative method in more detail.

As is well known, the Millian experimental inquiry is composed of four methods:

(i) Method of Agreement: comparing together different instances in which the phenomenon occurs.
(ii) Method of Difference: comparing instances in which the phenomenon does occur with instances in other respects similar in which it does not.

The first and the second can be combined in the Joint Method of Agreement and Difference, which consists in a double employment of the method of agreement, each proof being independent of the other, and corroborating it.

(iii) Method of Residues: subducting from any given phenomenon all the portions which can be assigned to known causes, the remainder will be the effect of the antecedents which had been overlooked or of which the effect was as yet an unknown quantity.
(iv) Method of Concomitant Variation, which deserves our special attention.

The Method of Concomitant Variation is particularly useful in case none of the previous methods (Agreement, Difference, Residues) is able to detect a *variation* of circumstances. For instance, in the presence of permanent causes or indestructible natural agents that are impossible either to exclude or to isolate, we can neither hinder them from being present nor contrive that they shall be present alone. But in such a case a comparison between concomitant *variations* will enable us to detect the causes.

Due to Mill's strong experimentalist view, the four methods are particularly well suited to the natural sciences. Interestingly enough, Mill (1843: ch. VI–VII, Book VI) argues that the experimental method is inapplicable in the social sciences. But even more interestingly, as we shall see next, Emile Durkheim (1895: ch. VI) vehemently goes against this view. Specifically, Durkheim claims that the method of concomitant variation is fruitfully used in sociology.

Emile Durkheim. In his masterpiece, Durkheim is interested in suicide as a social phenomenon. In *Le suicide*, he searches for the social causes of suicide, namely for the factors depending on which the social rate of suicide *varies* (Durkheim 1897: 332). Durkheim makes this very same claim at different places. In the Introduction, Durkheim (1897: 15) informs the reader that he intends to study the *variability* of the suicide rate[4] across time. This variability appears to be quite irrelevant across time within the *same* population, but is instead considerable across different societies. By examining how the suicide ratio *varies* across societies, Durkheim aims at detecting the social factors this variation depends on. Thus, the rationale of variation pervades also Durkheim's causal reasoning. For instance, Durkheim (1897: ch. II, Book IV) investigates the hypothesis of a relation between the nature of suicide and the kind of death chosen by the person who commits suicide. He argues

[4] The suicide rate is defined as the ratio between the number of voluntary deaths to the size of the overall population.

that the social causes determine the choice of committing suicide, but do not determine, *per se*, the type of suicide. He observes that the relative frequency of different types of suicide stays for a long time invariant within the same society but *varies* across different societies. Consequently, the reasons determining that choice—says Durkheim—are of a different nature: these are rather contingent conditions in which the person who commits suicide happens to live in. The kind of death chosen and the nature of suicide both depend on the social causes of suicide.

In *Le suicide* the rationale of variation can only be extrapolated from his argumentation along the essay. However, the rationale is definitively explicit in *Les règles de la méthode sociologique*. Chapter 6 opens thus:

> There is only one way of proving that a phenomenon is a cause of another, that is comparing cases where they are simultaneously present or absent, and looking whether the *variations* they exhibit in these different combinations of circumstances prove that one depend on the other. (My translation[5] and emphasis.)

The comparative method is, according to Durkheim the only way to make sociology scientific. However, not all procedures of the comparative method will do. Only the Millian Method of Concomitant Variations will. Witness Durkheim (1895: 160–161):

> In fact, for it [the method of concomitant variations] to be conclusive, it is not necessary that all variations that are different from the one we are comparing be rigorously excluded. A simple parallelism of values of the two phenomena, provided that it has been established in a sufficient number of cases sufficiently diverse, is the proof that a relation between them exists. (My translation.[6])

It is worth noting that Durkheim has a determinist conception of the causal relation, for he believes in the principle 'same cause, same effect'; the comparative method is scientific, i.e. conformed to the principle of causality, only if comparisons are analysed under the supposition that to the same effect always corresponds the same cause (Durkheim 1985: 157). In Durkheim a probabilistic characterisation of causal relations is totally absent. In case one of the two phenomena occurs without the other, this might be due, according to Durkheim, to the effect of a third phenomenon operating against the cause or to the fact that the cause is present in a different form.

Needless to say, this sounds like a problem of screening-off. Durkheim is of course aware of the fact that a concomitant variation might be due to a third phenomenon acting as a common cause or as an intervening factor between the first and the second. Such situations of possible confounding lead him to draw the following conclusion: the results of the method of concomitant variation have to be

[5] In the original: "Nous n'avons qu'un moyen de démontrer qu'un phénomène est cause d'un autre, c'est de comparer les cas où ils sont simultanément présents ou absents et de chercher si les *variations* qu'ils présentent dans ces différentes combinaisons de circonstances témoignent que l'un dépend de l'autre". (My emphasis.)

[6] In the original: "En effet, pour qu'elle [la méthode de variations concomitantes] soit démonstrative, il n'est pas nécessaire que toutes les variations différentes de celle que l'on compare aient été rigoureusement exclues. Le simple parallélisme des valeurs par lesquelles passent les deux phénomènes, pourvu qu'il ait été établi dans un nombre suffisant de cas suffisamment variés, est la preuve qu'il existe entre eux une relation".

interpreted. If a direct link from the cause to the effect is not self-evident, then the mechanism responsible for the concomitant variation has to be unveiled in order to rule out a case of common cause. For instance, says Durkheim, tendency to suicide varies as the tendency to schooling does. However, there is no direct causal link from schooling to suicide. Instead, this is a case of common cause: weakening of religious traditionalism strengthens at the same time the need of knowledge and the penchant to suicide.

Interestingly, in Durkheim (1895: 165) we can also find an 'invariance condition' *ante litteram*. What we have to compare is not just isolated variations but *series* of variations, regularly constituted. Durkheim (1895: 166–171) gives us other requirements these series have to satisfy which we do not need to examine in detail in this place. The main purpose has already been attained: the variation rationale also pervades Durkheim's methodology.

Foundations of epidemiology. The rationale of variation is further recognizable in more recent foundational texts in epidemiology. In *Causal thinking in the health sciences*, Susser (1973: 74 ff.) makes quite clear that to infer causal relations we need two fundamental elements: temporal asymmetry and comparisons between exposed and unexposed groups or cases and controls. For instance, Susser (1973: 65) claims that:

> Epidemiologists in search for causes want to make *asymmetrical* [emphasis in the original] statements that have direction. They seek to establish that an independent variable *X* causes *changes* [my emphasis] in the dependent variable *Y* and not the reverse […].

Here Susser is making a point about asymmetry. In epidemiology—and, let me add, more generally in the social sciences—time direction is an essential feature, unlike physics, where a symmetrical equation such as $F = MA$ (force equals mass times acceleration) can still have causal meaning. Later on Susser (1973: 81) says:

> The central problem of cohort studies, on the other hand, is to cope with the change that occurs with the passage of time. The study of cause involves the detection of *change* in a dependent variable produced by *change* in a independent variable. (My emphasis.)

The key point is to establish what variations are causal, not what regularities are causal. Similarly, in *Foundations of epidemiology*, Lilienfeld and Stolley appeal to *variational* evidence, i.e. evidence about the change in the frequency of disease a given factor is responsible for and not to regularity or invariance. As they put it (Lilienfeld and Stolley 1994: 263):

> In practice, a relationship is considered causal whenever evidence indicates that the factors form part of the complex circumstances which increase the probability of occurrence of disease and that a diminution of one or more of these factors decreases the frequency of disease.

In a similar vein, Lagiou et al. (2005: 565) claim:

> A factor is a cause of a certain disease when alterations in the frequency or intensity of this factor, without concomitant alterations in any other factor, are followed by changes in the frequency of occurrence of the disease, after the passage of a certain time period.

It seems apparent that the rationale of causality here involved is a rationale of variation, not of regularity nor of invariance.

4.1.3 Arguments from the Philosophical Literature

Let us now look at the philosophical literature. The intent is to show that many philosophers employ or presuppose this notion of variation, although they never make such claims explicit. It is commonly agreed that causes make effects happen. This is what Cartwright's (1989) capacities do, or, according to Woodward (2003), this is why causes explain their effects. Let me add, to cope with the problem of negative causes, that causes may also prevent effects from happening. In probabilistic terms, either causes raise the probability of the effect, or they lower it (in which case they are customarily called preventatives). In either case, however, we are interested in the *change* the putative cause is responsible for. As just stated, this is exactly the scheme of reasoning of probabilistic theories of causality, covariance structure models, structural equations models and contingency tables.

For instance, when Paul Humphreys discusses quantitative approaches to causality, he seems to propose this very same idea, namely that causality is tested by measuring variations. A contributing cause—he says—is a factor that "[…] contributes a precisely characterised quantitative amount to an equally precise quantified effect variable" (Humphreys 1989: 28).[7] Now, how is this contribution quantified? A unit change in a variable X_i caused a change in β units in Y, when all other factors X_j are held constant. Afterwards, Humphreys states the conditions this variation ought to satisfy. He mentions several, in particular, this change of β units must occur irrespective of the particular level other X_j factors happen to be on. That is, we require the structural parameter to be statistically stable. In this stability requirement, the experienced reader will have already recognized the celebrated *invariance condition*. Although the invariance condition is not the specific issue at stake for the moment, it comes into range at the right moment, for it leaves me room to stress an important matter. To provide a rationale of causality is not tantamount to suggesting a new condition to impose on causality, and it is not to depreciate the invariance condition either. On the contrary, the rationale of variation *conceptually* precedes the invariance condition, since what is supposed to be invariant is exactly the detected variation. I beg the reader to be patient, for the full argument in favour of conceptual precedence of variation will be given later.

This rationale emerges again in the discussion of the problem of empirical validity, when different models are compared. Witness Dan Hausman. In *Causal asymmetries* (1998: 224–225) he suggests considering two systems of linear equations (see systems 1 and 2).

[7] A contributing cause, in Humphreys' terminology, is a cause that raises the probability of the effect. In Suppes' words, a contributing cause would be a positive cause.

System 1	**System 2**
$x = \alpha_x + \mu_x$	$x = \alpha_x + \mu_x$
$y = \alpha_x y_x + \mu_y$	$y = \alpha_x y_x + \mu_y$
$z = \alpha_x z_x + \alpha_y z_y + \mu_z$	$z = \alpha^*{}_x z_x + \mu^*{}_z$.

> If $\alpha^*{}_x z = \alpha_x z + \alpha_y z \alpha_{xy}$ and $\mu^* z = \mu_z + \alpha_{yz}\mu_y$, then these two sets of equations are empirically equivalent. Data on values of *x, y*, and *z* are compatible with system 1 if and only if they are compatible with system 2. But if variables on the right-hand side of an equation are taken as causes of variables on the left-hand side, the two systems differ in what they say about the causal structure. […] The α's (including α^*) in systems 1 and 2 are the 'nonzero elements of the coefficient matrix' that the experimenters can choose. When the experimenters choose a new value for α_{xy} – suppose it is twice the old value – then if the error terms have zero mean and their distribution *does not change*, system 1 implies that the expected value of *y* will double and the expected value of *z* will increase by $2\alpha_{yz}$, while system 2 in contrast implies that the expected value of *z* will not change. If the expected value of *z* does not in fact change, then system 2 is correct. If it increases by $2\alpha_{yz}$, then system 1 is correct. (My emphasis.)

Hausman needs this argument to point out two presuppositions: first, the presupposition about the distribution of error terms, and second, the presupposition about parameter independence. Nonetheless, I would stress a third presupposition that he neglects. To claim which system, out of the two above, correctly describes the causal structure, experimenters employ a particular notion of causality, notably, as a measure of change. In fact experimenters will choose either system 1 or system 2 depending on which variations hold.

Listen to Hausman again. Later on, he comments on manipulability approaches and, in particular, on Woodward's; then he defines the asymmetry of modal invariance (Hausman 1988: 228):

> *X* causes *Y* if and only if one can calculate correctly (or make the best estimate of) what the value of *Y* would be if the value of *X* were set via intervention. (My notation.)

Again, I will not linger on the concept of invariance in this section. Hausman's definition of modal invariance deserves a note, though. His definition has at least two presuppositions. The first is the quite common assumption of the temporal asymmetry of causation. Because the whole book is about asymmetries, surely Hausman is aware of this one. The second, however, escaped his attention. Modal invariance presupposes that intervening on the putative cause *X* will produce a *change* on the putative effect *Y*, this variation being possible to compute, or at least possible to estimate.

In his latest book, Woodward (2003: 245–246). characterises causal generalisations as '*change*-relating' or '*variation*-relating'. This apparently presupposes a rationale of causality as measure of variation. Needless to say, the problem is that any generalisation that describes a correlation will be change-relating. So how can we distinguish genuine change-relating generalisations from spurious ones? The issue is indeed an epistemological one, and my answer will be sketched later in section 4.3. Woodward's answer, instead, is that genuine change-relating generalisation will show a certain invariability as prescribed by invariance. However, as I will argue at length later, the very notion of variation is presupposed by the condition of invariance. Moreover, differences and relations between the notion of variation and

the invariance condition will be examined further in the analysis of the features of causal models, in section 4.3.2.

4.1.4 Objections, or Further Evidence for Variation

Independence is more basic. At this point the reader might have the feeling that *variation* means something stronger, namely *dependence*. If so, Pearl argues instead that *in*dependence is more basic. For instance, Pearl (1988a: 385) claims that:

> […] conditional independence is not a 'restrictive assumption' made for mathematical convenience; nor it is an occasional grace of nature for which we must passively wait. Rather, it is a mental construct that we actively create, a psychological necessity that our culture labours to satisfy.

In other words, independence is an essential feature for causality. Nonetheless, a few pages later, Pearl seems to hold quite a different view, when he draws some conclusions about causal poly trees. He addresses the old question: causation or covariation? According to him, the threshold is in the notion of *control*: causal directionality between X and Y can only be tested through the introduction of a third variable Z. This is because, says Pearl (1988a: 397), by introducing Z we test whether

> […] by activating Z we can create *variations* Y and none in X, or alternatively, if *variations* in Z are accompanied by *variations* in X while Y remains unaltered. […] the construct of causality is merely a tentative, expedient device for encoding complex structures of dependencies in the closed world of a predefined set of variables. It serves to highlight useful independencies at a given level of abstraction, but *causal relationships undergo change* upon the introduction of new variable. (My emphasis.)

In the ultimate analysis, Pearl does use the rationale of variation, and this rationale seems to precede the notion of independence, contrary to what he himself states, i.e. that independence is the basic notion for causal reasoning.

Variation is a reformulation of regularist accounts. Another objection to my proposal is that this rationale is nothing but a reformulation of Humean regularist accounts. This is only partly true. Let me deal with the non-true part first. As is well known, Hume believed that causality lies in the constant conjunction of causes and effects. In the *Treatise of human nature* Hume says that, in spite of the impossibility of providing rational foundations for the existence of objects, space, or causal relations, to believe in their existence is a 'built in' habit of human nature. In particular, belief in causal relations is granted by experience.

For Hume, simple impressions always precede simple ideas in our mind, and by introspective experience we also know that simple impressions are always *associated* with simple ideas. Simple ideas are then combined in order to form complex ideas. This is possible thanks to imagination, which is a normative principle that allows us to order complex ideas according to (i) resemblance, (ii) contiguity in space and time, and (iii) causality. Of the three, causation is the only principle that takes

us beyond the evidence of our memory and senses. It establishes a link or connection between past and present experiences with events that we predict or explain, so that all reasoning concerning matters of fact seems to be founded on the relation of cause and effect. The causal connection is thus part of a principle of association that operates in our mind. Regular successions of impressions are followed by regular successions of simple ideas, and then imagination orders and conceptualizes successions of simple ideas into complex ideas, thus giving birth to causal relations. The famed problem is that regular successions so established by experience clearly lack the necessity we would require for causal successions. Hume's solution is that if causal relations cannot be established a priori, then they must be grounded in our experience, in particular, in our psychological habit of witnessing effects that regularly follow causes in time and space.

The crucial step in Hume's argument is significantly different from the rationale I propose. My claim is that we look for variations, not for regularities. Once the variation is detected, a condition of invariance or structural stability (among others) is imposed on it. What does structural stability give us? Not logical/nomic necessity, nor mere constant conjunction. Invariance recalls regularity; but that is profoundly incorrect. Structural stability is a condition imposed on the causal relation, in order to ensure that the model correctly specifies the data generating process and to ensure that the model does not confuse accidental relations with causal ones. Although the invariance condition is the most important one, other conditions grant the causal interpretation of statistical models. I shall carefully deal with this issue in section 4.3. Let me emphasise that here—i.e. in the discussion of the constraints to be imposed on causal relations—dwells the main reason why probabilistic theories cannot provide a meaningful rationale of causality. As stated in Chapter 2, the oversimplified formal structure of probabilistic theories inevitably leads to neglecting the multivariate aspect of social causality. Additionally, for the same reason they do not allow us to make all the assumptions nor the formal requirements needed for causal relations explicit and, consequently, do not allow us to make the rationale of causality explicit.

The true side of the objection is that in observational studies the attention is mainly directed to variations that happen to be regular, at least regular enough not to be occasional occurrences. But this is not to affirm *ipso facto* that those regular variations are causal, nor is it to state that those regular variations do not inform us enough to dare advancing a causal interpretation.

Before discussing the next objection, let me underline once more why so much importance is bestowed on the notion of variation. As stated above, variation is conceptually a *pre*condition. If causality is not set in the notion of variation, then it will be lodged in the invariance condition, which is conceptually misleading. The same holds for regularity. In both cases there is a further question to ask: invariance *of what*? regularity *of what*? The answer is, in both cases, *of a variation*. In particular, invariance—the queen of the causal conditions—only makes sense *within* a causal model, whereas variation is exactly what motivates testing invariance. In other words, without variation invariance is devoid of meaning. This is why variation conceptually precedes invariance. To provide a rationale of causality means, to

put it otherwise, to give the bottom-line concept. Neither invariance nor regularity are apt to accomplish that task. But variation is.

Hume inferred causation from regularity, whereas my claim is that we infer causation from variation because variation conceptually and empirically comes before regularity. Of course, both notions—regularity and variation—do not guarantee a straight causal interpretation, but the rationale of variation puts us on the right track because it makes causality an empirical issue rather than a psychological fact or a mere reduction to statistical conditions to be satisfied. Also, Hume's regularity view immediately rules out singular causation, i.e. causal relations that occur only once and that therefore do not instantiate regularities. Instead, the variation rationale is still meaningful in the case of singular causation, as I will show in section 6 where I discuss Ducasse's account (Ducasse 1926).

The Humean paradigm of regularity still dominates philosophy and methodology of science. The regularity views of causation are clearly a heritage of the Humean account. Defenders of the regularity view claim, roughly, that to assert a causal relation between two events x and y means to assert the existence of a law such that every time an event x of type X occurs, then an event y of type Y will invariably follow. Humeans like Mill or Mackie have advanced more sophisticated versions of the regularist view (e.g. the I.N.U.S. condition) and tried to characterise the kind of regularity that can underpin causal relations by tying causation to laws of nature. The same program was initiated by Carnap and other logical empiricists. Even Lewis' counterfactual approach uses regularities as means to capture the conditions under which counterfactual assertions are true. Similarly, present-day regression techniques put emphasis on average causal effects, i.e. factors that operate regularly and invariably enough to be deemed causal. Also, the probability raising requirement often advocated in probabilistic accounts, has been backed by a ceteris paribus condition—i.e. a condition of homogeneity—that makes things regular enough to let the cause raise the probability of the effect.

The variation rationale profoundly breaks down this received view. The variation rationale claims, in the first place, that this emphasis on regularity is not well-founded, for regularities themselves require a prior notion, which is variation. Secondly, the difficulty or impossibility to establish causal laws in the social sciences is usually taken as a structural weakness or even as an intrinsic impossibility for the social sciences to reach the kingdom of hard sciences. The variation rationale is a first step for a radical change in the dominant paradigm: if, in the ultimate analysis, causal modelling aims at measuring variations rather than establishing regularities, this might be due to the fact that the regularist rationale is not, after all, as well founded as empiricists have claimed since Hume. This calls for a change of paradigm in causal modelling, rather than giving up our endeavour to establish causal claims or questioning the rigorous scientific character of the social sciences.

We cannot find variation in homogeneous populations. Another threat to the variation rationale is when the population is highly homogeneous. The objection has been raised in conversation by Professor Wunsch. Here is his argument. Suppose we are interested in the social conditions of women in Islamic countries: let our causal hypothesis be that Islamic culture (broadly understood) has causal

effectiveness on gender equality. It is very likely that no variation will be detected if data is gathered from an Islamic country, the reason being exactly that such a population would be too homogeneous to detect a variation. Hence, one might think that this constitutes a crucial objection to the rationale of variation. I shall argue that this is not the case. Homogeneity surely is a *methodological* threat that practising scientists have to deal with, but it is not a knock-out objection against the epistemology here proposed. Let me go over the argument—I shall take advantage of the scheme of reasoning underlying Salmon's S-R model of explanation. In doing this, I shall favour an oversimplified exposition—not trivial, though—to better grasp the intuitions behind my defence.

Let us consider the two case studies presented in Chapter 1, namely smoking and lung cancer and mother's education and child survival. By way of a reminder, Peto et al. have tried to assess the risks of smoking and of smoking cessation in the UK population since 1950. In the 1990 study relative risks for men and women are calculated comparing categories of smokers with lifelong non-smokers. So, roughly speaking, the risk of lung cancer is tested over the partition {smokers, non-smokers}. Relative risks are then combined with national lung cancer mortality rates to estimate the absolute hazards in various categories: smokers, former smokers, and non-smokers. Again, to evaluate the risk of lung cancer, the population is ideally partitioned thus: {smokers, former smokers, non-smokers}. On the other hand, Caldwell has shown that mother's education is a crucial causal factor for child survival, not reducible to other socio-economic factors. That is to say, no other factor has the impact that maternal education has, and in their totality those factors do not even come close to explaining the effect that maternal education has. Causal effectiveness of education is thus assessed by measuring the proportion of children dead over the partition {no schooling, primary school only, at least some secondary school}.

The strategy of partitioning the reference class is very much akin to the basic idea of probabilistic theories of causality, namely we work out the difference between the probability of the effect when the cause is present and when it is not: $P(E|C) \stackrel{?}{\neq} P(E|\bar{C})$.

If the putative causal factor *C does* make a difference, then it will be considered in the explanation and neglected otherwise. To recall, according to Salmon's S-R model, an explanandum-partition is imposed upon the initial reference class and cells in the explanans-partition must be (objectively) homogeneous, or otherwise it would mean that a relevant explanatory factor has been neglected. The point I want to make through these unsophisticated S-R explanations is the following: the rationale of variation is feasible because a partition of the population exists—in Salmon's words, because the explanandum-partition exists.

This is patently not true in Professor Wunsch's example. If we want to assess the causal impact of Islamic culture on gender equality, the partition of the reference class {Islamic, non-Islamic} will not do, simply because this partition does not exist in a homogeneous population where virtually everyone is Muslim. Intuitively, to see whether Islamic culture brings about differences in gender equality, we would need to broaden the reference class so that the partition {Islamic, non-Islamic} exists.

These are just speculations over a fictitious example, though. Let us then consider a real case study.

In 1996, Samuel Huntington[8] states the 'clash of civilization' thesis: the end of the Cold War brought new dangers, and future conflicts will take place between people belonging to different *cultural* entities. The line that formerly divided Europe has now been moved further east, dividing Western Christianity from Muslim and Orthodox people, the threshold being *democracy*. This influential account seemed to offer insights to explain the causes of violent ethno-religious conflicts, for instance in Bosnia, the Caucasus, or in the Middle East. But the clash of civilization thesis might possibly be used to explain the root causes of the events of September 11th.

Norris and Inglehart, in their 2003 paper, examine an extensive body of empirical evidence relating to this debate. In a nutshell, here is their result. The clash of civilization thesis is right in claiming that culture matters and, indeed, it matters a lot. Nonetheless, Huntington is mistaken in thinking that the core clash between the West and Islamic countries concerns democracy, for many non-Islamic countries bestow the same importance on the role of religious leadership as Islamic ones do. Moreover, the clash of civilization thesis fails to identify the most basic difference between the West and Islam, which lies in *gender equality and sexual liberalization*. And this is why this case study is of great interest here.

That study examines cultural values across seventy-five countries, including nine predominantly Islamic societies. To test the evidence for the clash of civilization thesis, this study compares values at the societal-level, based on the assumption that predominant cultures exert a broad and diffuse influence upon all people living under them. To analyse the survey evidence, societies are classified into categories based on the predominant religious identities within each nation. Here is the partition: {Protestant, Catholic, Islamic, Orthodox, Central Europe, Latin America, Sinic/Confucian, Sub-Saharan African}. The comparative evidence suggests that there is a *substantial* cultural cleavage in social beliefs about gender equality and sexual liberalization. Differently put, to assess whether Islamic cultures affects gender equality is possible thanks to a very broad initial reference class, where an explanandum-partition exists. Given this partition, differences in the cells can be compared and the rationale of variation works fine again. To sum up, the *epistemological* point still holds: causality is tested by measuring variations. However, the threat of homogeneity does not go away without leaving any traces behind. Instead, we are left with a very important lesson. In fact, at the *methodological* level, one clause should be added: causality is tested by measuring variations in a suitable reference class where an explanandum-partition exists.

Nevertheless, my take-home message does not completely persuade Professor Wunsch. Here is his reply. *First*, widening the research domain could render causality highly context-dependent, notably by interaction effects with other variables. *Second*, it might be the case that it is not possible to broaden the research domain because a given characteristic is not *de facto* present somewhere else. So—he

[8] Huntington, S. P. (1996). *The clash of civilization and the remaking of world order*. New York: Simon and Schuster. Quoted in Norris and Inglehart (2003).

argues—the objection still holds; according to him, a possible solution would be to control covariates and compare the population under study with other populations.

Although I understand Professor Wunsch's point, I still wonder: what would controlling the covariates be useful for? Controlling the covariates would let a *variation* emerge. On the other hand, comparing results that refer to different populations and come from different studies, opens a large debate. Whether different studies be comparable at all, or whether results from one study might be applicable to other populations is in fact a serious challenge to the external validity of causal models. But this does not undermine my epistemological claim. This would rather show that, in fact, the rationale of causality social scientists have in mind is this rationale of variation. And with this epistemological awareness we have to work out feasible solutions to methodological threats, such as homogeneity.

Variation is not involved in the causes of states. Finally, an interesting question arises when we consider the causes of *states*. For instance, suppose we ask why a boat floats, that is, we wonder what the causes of the state of the boat are. In such cases no variation seems to be involved, at least at first glance. The objection may be addressed in several ways. To begin with, this simple example concerns the natural sciences rather than the social sciences, and therefore might fall outside our scope. However, questions about the causes of states concern the social domain too and therefore the objection becomes directly relevant to us.

The second thing worth noting is that questions about the causes of states are, in fact, quests for causal *explanations*. In the philosophical literature, various accounts of causal explanation have been offered. Salmon (1984), for instance, maintained that we need to provide the physical (causal) process. By definition, a process involves states that come one after another. Thus, it seems that an explanation in terms of processes will not do in this case, for what we are seeking to explain is just *one* state in the process. Yet, one might rebut this by saying that a specific state is what it is because of the spatio-temporally prior state. However, in this specific case, it seems that we are not just concerned with the state of the boat at a particular point of space–time, for which a physical (causal) process could be provided. Instead, what we are looking for is something more general—that is why, under certain circumstances, boats float. In other words, here we are looking for the *laws* that regiment the phenomenon.

A boat floats because some relevant physical laws hold, for instance Archimedes' law. Traditionally, Hempel and Oppenheim (1948) offered a covering-law model that emphasises the explanatory role of laws. In this case, we would have to state the initial conditions, the laws, and from those premises we could deductively infer the phenomenon to be explained, i.e. that boats float. The question now is whether variation is involved in this kind of explanation, and if so, how. At first sight, variation is not involved in the explanation of states, unless we modify the original question slightly, for instance by asking why the boat floats rather than, say, sinks. Then, the answer might involve invoking a *variation* in the initial conditions, for instance the ratio between empty and fraught parts of the boat. Another possibility to explain this phenomenon would be to appeal to van Fraassen's pragmatic approach (van Fraassen 1980). Simply put, an explanation is an answer to a particular

why-question. In this case, the why-question concerning the boat floating will involve evaluating a contrast class (i.e., floating rather than sinking, or flying, etc.) and the explanation will be constituted by an answer linking the explanandum to a set of possible explanans. A good answer—i.e., a good explanation—will thus require us to specify a relevance relation between the question and the answer. In this case, the relevance relation between the explanandum and the explanans will be a nomic and causal relation.

But let us focus on laws. Interestingly, a variational reading of laws can be offered: laws describe how things change in time. As Maudlin (2007: 11–12). says:

> What is most obvious about these laws is that they describe how the physical state of a system or particle evolves through in time. These laws are generally presumed to be universally valid. But this is not a feature directly represented the formulae, nor does it appear to be essential to their status as laws. […] The laws cited above [i.e. dynamic laws], then, tell us how, at least for some period and in some region, physical states evolve through time.

According to Maudlin, the variational character of law statements is more important than the fact that they are (supposedly) universally valid. Let me clarify this by saying that what is implied by this 'variational reading of laws' is that the object of the law varies. So, if we credit this variational character of laws, then, given that we need to invoke laws in the quest for the causes of states, it follows that variation *is* involved in such cases. However, Maudlin does not push this idea further, and the change of paradigm his view might engender rests, in his account, a possibility rather than a real alternative to the received view on laws. Unfortunately, to develop a variational account of laws goes far beyond the scope of this work and I will leave it as a path for future research, though it is—it seems to me—a very promising one.

The third thing worth noting is that when we are concerned with the causes of states and we offer a covering-law explanation, we presuppose that the relevant physical laws hold—we do not question *how we came to know* about those laws. If we were to answer *this last* question, then we would have to investigate whether or not the rationale of variation is at work. Again, this concerns the natural sciences and thus goes beyond the scope of the present work. However, if we grant, as Maudlin does, that the "most obvious" thing about dynamic laws is that they describe how things change in time, then the rationale of variation will be also at work here, for it is through the observation and experimentation of the *variation* of the behaviour of the object that dynamic laws are construed. The regularity of the behaviour is, again, a constraint we impose on the variation in order to claim its generality.

Instead, let us consider an analogue example in the social domain. Why is divorce rate so high nowadays? Or, why does school failure rate increase? Given that, arguably, there are no laws in social science, or at least they are not so easily detectable (see also section 3.3 on this point), we can hardly offer a covering-law explanation. If we cannot appeal to laws, what else might explain states? Here, what we have to provide is the *mechanism* underlying the phenomenon at hand. A thorough discussion of mechanisms and their modelling will be offered later in section 6.1, so I will just anticipate the main ideas. Consider the case of high divorce rates. What we need to find out are the causes of divorce rate and how these causes interact

with each other in order to deliver the rate we observe. It is in this *seeking for the causes* that the rationale of variation is involved. The reason is simply that 'seeking for causes' requires causal modelling, and, in turn, causal modelling is regimented by the rationale of variation, as shown earlier in section 4.1. Moreover, as we shall see in section 6.1 in more detail, mechanisms are also modelled according to what *variational* relations between variables hold.

A last thought before closing this section. *Why do boats float? Why is divorce rate high?* Are those questions the same? Although we might agree that both are questions about states, surely there is a difference between them. Whilst we assume that (ceteris paribus) boats always floated and will always do, divorce rate is not necessarily (almost certainly not) the same across time and space. Therefore, in the first case we look for an 'a-temporal' explanation and for a 'time–space' dependent explanation in the second. Therefore, in explaining the causes of states of a social phenomenon, the primary interest is in what makes the phenomenon what it is, as opposed to what it was or as opposed to what it will be after an intervention. In other words, an interest in the causes of social states is an interest in the *variation* with respect to previous or future states.

4.2 Varieties of Variations

Up to now, the notion of variation seems to have provided a meaningful rationale of causality in causal modelling. However, the presentation given in section 4.1 is general in scope. In practice, we may look for different types of variations, depending on the case at hand. A taxonomy of variations can be sketched according to five different criteria, or taxa:

(1) Variations across time.
(2) Variations across individuals.
(3) Variations across characteristics.
(4) Counterfactual and control group variations.
(5) Observational vs. interventional variations.

The case studies analysed in Chapter 1 will help to illustrate these five taxa.

In the health and wealth case study, variation is about the same individuals but at different times. Health and wealth histories are in fact analysed for the same sample in three wave surveys (taxon 1). Also, the whole statistical set up is meant to examine whether past events have an influence on later events (taxon 2). Moreover, detected variations are observational, for observational data is used, and no experimental intervention is performed (taxon 4).

Let us now consider Caldwell's model of the impact of mother's education on child survival. In this case we are interested in how variations in a certain variable are tied to variations in another variable; namely, we are interested in variations among different characteristics (taxon 3). In this case time is not taken into account, because the focus is not on how the impact of mother's education varies across time (taxon 1), which would be the case by analysing longitudinal biographical data.

The difference with the previous case is apparent. In the health and wealth example, a sample of elderly Americans is interviewed, and data concerning their health and wealth histories is collected. For the same group of individuals variations from earlier health events to later wealth events (and vice-versa) are analysed. In Caldwell's model, two characteristics are particularly monitored across different individuals—in fact, data from different surveys of different populations is compared.

Consider now epidemiological studies on smoking and lung cancer. Let us see what kind of variation Peto et al. (2000) are interested in. Hospital patients under 75 years of age (with and without cancer) in 1950 and 1990 participated in this study. The goal is thus to relate trends in smoking and smoking cessation. We deal here with a variation across time (taxon 1) and across the same individuals (taxon 2). In fact, following the health history of the participants, researchers come to calculate different risks of death, depending on age and trends in smoking. Results come from observational data; thus this is again an example of observational variation (taxon 5).

To illustrate taxon 4, consider studies where two different groups—a test group and a control group—are involved. In this case, variations are quite similar to counterfactual variations used, for instance, in policy prediction. In taxon 4, counterfactual is not intended in Lewis' sense (Lewis 1973), namely the antecedent of the subjunctive conditional is known to be false. Counterfactuals are here understood *à la* Pearl (2000: 201 ff.), that is, counterfactuals underlie the typical scientific reasoning in policy predictions and decision making. It is worth noting that Pearl's account escapes the objection of hyperrealism, for it is not based on a possible-world semantics but on the apparatus of causal models. I will first explain what counterfactual variations are, and then explain why test and control group variations are of that kind. I shall also discuss both types of counterfactual accounts in relation to the rationale of variation later in section 7.2.

Consider again the typical economic example of demand and price equilibrium presented at the beginning of Chapter 3. We had a system of two equations; in the first equation the quantity of household demand for a certain product is the function of the unit price of the product and the household income (plus errors); in the second equation the unit price of the product is the function of the quantity of household demand and the wage rate for producing the product (plus errors).

A counterfactual question would ask, for instance, given the current price of a certain product, what would be the expected value of the demand, if we were to control, i.e. to set a different value for the price of the product? So, given the current price of the product, we wish to calculate the expected value of the demand, had the price changed from x to x_1.

In the case of test and control group studies, we go beyond initial hypothetical conditions. Different conditions are *de facto* set up as initial conditions or as interventions on them, and results are subsequently compared. It is in this sense that different set ups of initial conditions give the counterfactual character of the variation. Another typical example is the evaluation of the effectiveness of a medical treatment, compared to the absence of the treatment or to the administration of a placebo. This would easily lead us into never-ending discussions about the role of

randomization or possible confounders due to the placebo effect. Worrall (2002) and Cartwright (2007b) brilliantly present the *status quaestionis* about evidence-based medicine and randomization, and also give a rather complete list of references, well informed both from the philosophical and scientific literature.

Counterfactuals and control group variations also help in clarifying taxon 5. Although in the social sciences we are more often concerned about observational variations, researchers attempt to qualify interventional variations requiring that variables be *intervenable*, at least in principle. These attempts to characterise interventions mainly belong to graphical methods for causal analysis, as for instance in Bayesian nets. As I mentioned in section 3.1.1, every graph is associated with a joint probability distribution of the variables in the graph. If we intervene on this graph, then the distribution should change accordingly. To intervene on a graph means, to put it in a very simple way, to add or to remove nodes, or to cut off arrows between nodes. When these modifications are causal, then the former joint distribution should change.

4.3 What Guarantees the Causal Interpretation?

The problem of the validity of causal inference, old as it may be, is still a matter for debate in causal modelling. Both philosophers and statisticians attempted to give an answer either relying on metaphysical arguments—true correlations *are* causal—or by requiring that more and more severe conditions be met. For instance, in *Causal asymmetries*, Dan Hausman makes the strong claim that there are no accidental probabilistic dependencies. This means that correlations prove causation, because genuine correlations are themselves causal. Conversely, accidental correlations, like the one due to the monotonic increase of both bread prices in England and sea levels in Venice, are simply not true correlations—that is, we picked up the wrong correlation, so to speak. Hausman (1998) makes similar claims throughout the book (for instance, Hausman 1998: 33, 56). On the other hand, as Pearl (2000: 133) correctly notices, the prevailing interpretation of SEM nowadays differs from the one intended by its originators. Structural equations are often interpreted as simple carriers of probabilistic information; they may carry information about density function, or summarise information conveyed by covariance matrices, but the causal interpretation is hardly allowed.

Moreover, although some authors do emphasize the pivotal role of assumptions in causal modelling, the investigation is not deep enough. For instance, Freedman (2004a) distinguishes two types of assumptions—statistical and causal—and acknowledges that without interventions causal inference is problematic. However, the causal assumptions he has in mind consist of *assuming* that structural equations unveil the causal mechanism that generates the observational data. It seems to me that this is not just a causal assumption, but a metaphysical one. According to Holland (1986), associational models make descriptive claims about conditional distributions, whereas causal models, in addition, aim at evaluating statistical relevance relations in order to quantify the causal effect of a treatment or intervention.

Stone (1993), (partly) discloses what the causal assumptions are; he ranks them from the strongest (covariate sufficiency) to the weakest (ignorable treatment assignment) and nicely shows how, in the ultimate analysis, they are all implied by the strongest, namely covariate sufficiency. Nonetheless, Stone fails to mention that causal assumptions do not tell the whole story. The debate is still open.

As just stated, causal models are tested by measuring suitable variations. However, variations might be accidental, and to provide a promising rationale of causality does not mean that a causal interpretation of statistical results follows straightforwardly. Consequently, I have to face the problem of the validity of causal inference. So here is the question: when are we entitled to interpret variations causally? The answer to this question will be the subject matter of the present section.

The distinction between associational and causal models on the one hand, and a careful analysis of assumptions in both associational and causal models on the other hand, seems to me a favourable path, worth exploring further. Here is the intuition in a nutshell: assumptions and background knowledge in causal models play a major role—the causal interpretation comes from specific causal assumptions that simple associational models do not have. To support this claim, I shall provide both methodological and empirical arguments.

In the first place, I will examine models employed for associational purposes, and subsequently models employed for causal purposes. In particular, this scrutiny will show that the difference between associational models and causal models is located in their set of assumptions and in the use of background knowledge. In what follows, I will not linger on the technical aspects of statistical models, for a detailed presentation of the meaning of equations and assumptions has already been provided earlier in Chapter 3. Instead, the main focus here is to find a solution to the problem of the warranty of the causal interpretation.

4.3.1 Associational Models

Let us consider associational models first. In associational models, the goal is to describe how variables co-vary—i.e. to describe how a certain variable that we choose to be the dependent variable varies depending on other variables (the independent ones). A number of assumptions are normally made.

Features of Associational Models, or Statistical Assumptions

(a) Linearity (in the parameters and in the dependent variable).
(b) Normality.
(c) Non-measurement error.
(d) Non-correlation of error terms.

We suppose the model to be linear, or approximately so, and normal. This is for convenience. A linear model is easy to manipulate, easy to estimate statistically, and the resulting estimators have nice properties. Tests for linearity can be performed,

but linearity is often assumed. In case linearity fails, generalised linear models can be used instead. The same holds for normal distributions. In particular, normality is assumed for error terms, but not necessarily for other variables. We also assume that variables are measured without error, and that the errors are not correlated with the independent variables. When errors are not correlated, we obtain the best linear unbiased estimate of parameters.

Let us call this set of assumptions the *statistical assumptions*. When these assumptions are satisfied, the conditional distribution correctly describes how variables co-vary. Under statistical assumptions the model is identified and parameters estimated by least square techniques. Associational inference is thus statistical inference (estimates, tests, posterior distributions) about the associational parameters relating X and Y on the basis of data gathered about X and Y from units in the population. It is worth noting that variables in associational studies play a symmetrical role, with no variable designated as the effect, namely they are simply employed to study what relations hold between variables, but not to state any causal claim concerning those relations. Associational models are used, for instance, to study the relation between height and weight.

The moral is that if the statistical assumptions are satisfied, then the conditional distribution correctly describes co-variations between variables. However, at this stage, no causal interpretation is allowed yet. That is to say, there is no necessary causal information conveyed by associational parameters, nor is it generally valid to give the regression coefficients in the equations the causal interpretation. Let me emphasise that associational models typically do not use a hypothetico-deductive methodology. In fact, they are not designed to confirm or disconfirm any causal structure; rather, they are simply designed to explore relations among variables. To illustrate, here is an example.

A Case Study: Obesity and Educational Level. The *Journal of Epidemiology and Community Health* published a study on obesity and educational levels in Spain (Gutiérrez-Fisac et al. 2002). Three cross sectional studies, representative of the adult population in Spain, were carried out in 1987, 1995, and 1997. The objective was to determine the size of obesity differences associated with educational levels in the adult Spanish population. The statistical model used in this analysis is a binomial regression model that satisfies *just* the standard statistical assumptions described above. As a result, obesity prevalence is highest in people with elementary education. As the authors say in the conclusions of their article, social mobility probably explains the association between socio-economic variables and obesity; however, no causal claim follows directly from their analysis.

At this point the reader might be wondering whether there is any inconsistency in my variation rationale, for I claim that associational models aim at studying how variables co-*vary* and yet do not allow to interpret variations *causally*. I do not think there is. If the associational study detects interesting and meaningful variations, then further investigation is needed in order to establish a causal variation. The question is, of course, what makes the detected variations interesting and meaningful. The

answer is, in a nutshell, background knowledge and causal context. If no variation is detected, it does not make sense to continue with a second stage, that is with a causal model, which will be discussed next.

4.3.2 Causal Models

Let us now consider causal models. In causal models, the goal is to go beyond descriptive variational claims and evaluate statistical relevance relations or conditional distributions to quantify the causal effect of the putative cause(s). More generally, we can say that the goal is to test a given causal structure. To go beyond the variational descriptive claims made in associational models, we need, on the one hand, accurate knowledge of the causal context, and, on the other hand, supplementary assumptions in the model.

Features of Causal Models

(a) Knowledge of the causal context. E.g.: previous studies, available scientific theories, background knowledge. Background knowledge should, whenever possible, include whatever information makes the conceptual hypothesis plausible and worth submitting to empirical testing.

(b) Conceptual hypothesis. This is a causal claim about the hypothesised link between conceptual variables. The conceptual hypothesis states the hypothesised causal structure, and is put forward for empirical testing. The rationale of testing is given exactly by the rationale of variation discussed above in section 4.1.

(c) Extra-statistical assumptions (untested assumptions).

(a) Direction of time.
(b) Causal asymmetry.
(c) Causal priority.
(d) Causal ordering.
(e) Causal mechanism.
(f) Determinism.

Although the direction of time still represents a fashionable and controversial topic in the philosophy of physics, causal modelling normally assumes that the second law of thermodynamics gives the direction of time from the past to the future, in the same direction as the increase of entropy. Also, three characteristics of the causal relation are generally assumed. Causal relations are asymmetric; causes precede effects in time; and these two assumptions, together, guide the choice of causal ordering. That is to say, the temporal order in which variables are observed is assumed. According to causal asymmetry, if X causes Y, Y does not cause X. The problem of feedback loops is sidestepped by taking into account causal priority: $X_t \rightarrow Y_{t'} \rightarrow X_{t''}$, where $t < t' < t''$. That is, Y is considered as an intervening variable between two temporally distinct values of X.

Furthermore, in some cases (for instance, in the smoking and lung cancer study presented in section 1.2) statistical models are not the lieu where the causal mechanism is tested: that is, the *physical* mechanism is simply assumed. In other cases, however, the causal mechanism will be exactly the object of modelling—on this point see later section 6.1. Notice, however, that to assume the physical mechanism does not grant *ipso facto* the causal interpretation, as Freedman suggests. The issue of determinism is more debatable. However, the structural equation can be considered a deterministic function where probabilities come in through error terms that represent our lack of knowledge. Determinism has been presented as a *methodological* assumption, as opposed to *metaphysical* determinism held, for instance, by Pearl (2000). Pearl's determinism is nicely criticized by Williamson (2005b).

Let me emphasise that the aforementioned extra-statistical assumptions are *not* criteria to assess causality but (some) of the untested assumptions about the characteristics of causal relations. The difference between a criterion and an assumption could be explained as follows. Consider, for instance, causal priority. As an assumption, causal priority means that we take for granted that causes precede effects in time. As a criterion, instead, the task is to check whether the putative cause C precedes the putative effect E in time, and if the test is passed, this will be *evidence for* the causal relation between C and E. I will now go over a few case studies to illustrate these features.

A Case Study: Health System and Mortality. In 1992 the *European Journal of Population* published a causal analysis about the incidence of the health system on regional mortality in the adult Spanish population (López-Ríos et al. 1992). This paper is instructive because it clearly shows most of the features just mentioned. In particular, knowledge of the causal context is given by knowledge of the political, economic, and social situation in Spain over the decades 1970–1980, which led to a lower mortality rate at the time of the study. Knowledge of the socio-political context makes clear the modelling strategy of this study. In fact, previous studies in demography and medical geography examined the incidence of the health system on regional mortality and came to the conclusion that regional differences in mortality could not possibly be explained by regional differences in the health system. However, Spain met deep socio-economic changes in the mid-Seventies, and consequently policy in that period simultaneously tried to intervene by improving on the social and economic situation.

The causal model put forward in this study highly depends on this background knowledge; in fact, mortality is influenced by the health system which is in turn influenced by the social and economic development. It is this background that explains the choice of distinguishing the supply and demand of medical care, unlike the majority of similar ecological studies (see Fig. 4.5).

Concerning the extra-statistical assumptions, causal priority and causal ordering are needed. In fact, a temporal gap is assumed between the putative causes (health system, economic, and social development) and the putative effect (mortality).

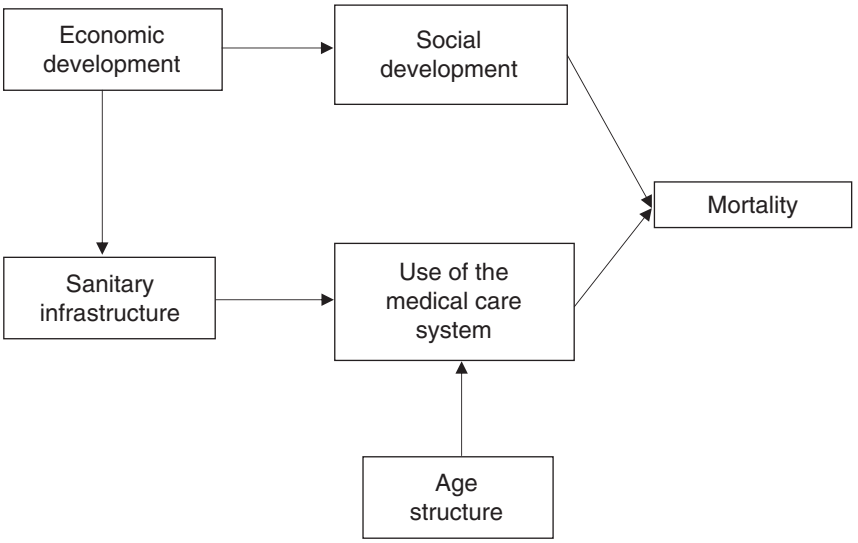

Fig. 4.5 Health system and mortality

A Case Study: Smoking and Lung Cancer. Instead, the causal mechanism (feature **e** above of the extra-statistical assumptions) is clearly assumed and not tested in one of the numerous studies on the causal impact of smoking on lung cancer (Peto et al. 2000). In fact, in this particular analysis the goal is to relate UK national trends since 1950, in smoking, smoking cessation, and in lung cancer. How cigarette smoking physically affects lungs and produces cancer is not addressed in this analysis, but rather assumed.

Let me now spend a few words on background knowledge. As I briefly mentioned in Chapter 1, there might be different sources of background knowledge: (i) similar evidence about the same putative mechanism, (ii) general knowledge about the socio-political context, (iii) knowledge of the physical-biological-physiological mechanism, (iv) use of similar or different methodologies or of data. Different studies normally consider different populations. Differences can accordingly concern time, geographic location, basic demographic characteristics, etc. So, when is the appeal to background knowledge justified and when is it not? Background knowledge can nonproblematically be used only if populations are in fact comparable. For instance, Adams et al. invoke similar evidence about the same causal mechanism detected in different studies. Caldwell invokes as background knowledge information about the socio-political context: it is because developing countries of West Africa are significantly different from developed countries that maternal education plays a causal role for child mortality in the former but not in the latter. In Peto et al.'s study, background knowledge is instead constituted by knowledge of a biological mechanism that explains why tobacco is a carcinogenic substance.

There is, however, a patent difference between the case of Peto et al. and the case of Caldwell. Background knowledge invoked by Peto et al. is the current theory of carcinogenesis: cancer occurs when genetic mutations disturb the normal regulation of a cell and erroneously cause it to multiply—with potentially fatal consequences. This theory today constitutes a *well-established* theory of carcinogenesis. But the background knowledge invoked by Caldwell is not at all a well-established theory, for it is exactly this theory he is trying to build. Background knowledge in Caldwell's study is constituted by much weaker similar conditions that make different studies comparable.

In section 4.1, I discussed a new methodology of research in quantitative demography. This methodology, called double-frontiers approach, is applied to analyse the status of the newborn at birth. Because it is the first time that this methodology has been applied in demography, the structural approach of the classical model of Mosley and Chen[9] is taken as background knowledge: if the double-frontiers approach gives results consistent with the classical model, then this will positively contribute to the soundness of results of the new approach. Similarly, Adams et al. devote an entire section to the discussion of the foundations of econometric causality tests and argue in favour of Granger-causality (with some refinements).

At this point I cannot refrain from asking: *background or foundations*? The difference between the two appears to be subtle but essential. It seems to me that in the case of the double-frontiers approach and in the Adams et al. study, different methodologies are considered as stronger than simple background. The structural model developed by Mosley and Chen on the one hand, and the econometric tests for causality developed by Granger on the other, are meant to provide justification for the results achieved (in the first case) and for the methodology to be employed (in the second).

I would like to underline the consequences of that. To mention Mosley and Chen's or Granger's approach as foundations means to take them for granted. It means, in other words, to consider these methodologies as well-established and the subsequent results of applications as reliable. Of course I am not criticising this practice in order to preach that every study in the social sciences ought to start by a deep questioning of the methodologies and of the previous results. But *this place* is the right one to do that.

Are Granger-causality tests actually capable of testing causality? Are structural relations in Mosley and Chen's model causal? Were the answers to be positive, then we could consider them as *foundations*. Were the answers to be negative or highly dubious, then the social science research would show need for philosophical help to establish better foundations for its methods. Let us now revert to the features of causal models.

(d) Statistical assumptions. The same standard statistical assumptions are valid for causal models too. This is tantamount to saying that causal models are

[9] The model for child survival in developing countries developed by Mosley and Chen (1984) will be thoroughly discussed in section 6.1, where the problem of modelling mechanisms is addressed.

augmented associational models, augmented precisely with features described in **(a), (b), (c)** above, and **(e)** below.

(e) Causal assumptions:

(a) Structure of the causal relation:
  (i) Linearity (as in associational models).
  (ii) Separability: the causal function is additively separable between the independent variables and the errors.

(b) Covariate sufficiency: the independent variables are causes of the dependent variable, and these are all the variables needed to account for the variation in the dependent variable.

(c) No-confounding: other factors liable to screen-off the causal variables are ruled out.

(d) Non-causality of error terms: there are no causal relations between the independent variables and errors.

(e) Stability: invariability of distributions.

(f) Invariance condition: invariability of the causal relation.

We are finally in a position to discuss the specific set of assumptions that make a causal model *causal*. Those assumptions ensure that correlation coefficients in the equations are in fact *causal* coefficients, and thus enable us to correctly specify the data generating process. Additional assumptions concern, for instance, linearity and additivity of the model. Linearity, strictly speaking, is not a necessary feature of causality. However, it is a useful assumption in practice. Of particular importance is the assumption of covariate sufficiency. This assumption conveys the idea that the independent variables are causes of the dependent variable, and that these are *all* the variables that are responsible for the variation in the dependent variable. The assumption of no-confounding, furthermore, means that we ruled out other factors liable to screen-off the variables we took into account. Are covariate sufficiency and no-confounding conditions we test or conditions we simply assume? To some extent these are genuine assumptions made at the choice of variables stage and that crucially depend on available background knowledge and on the use we make of it. Besides, these assumptions are crucially causal because they fix which variables play the role of causes and which ones do not.

The difference between stability and invariance is more subtle. On the one hand, stability of distributions is assumed to ensure that variations in the parameterizations do not destroy independencies between variables, viz. conditional independencies in the causal model are said to be stable when they are invariant under parametric changes. Pearl (2000: 48) calls this condition 'stability', but stability is elsewhere called 'DAG-isomorphism' (Pearl 1988a: 128), or 'faithfulness' (Spirtes et al. 1993). Let me explain further. There are two types of independencies: those that would persist for every choice of parameters, and those that are sensitive to the precise values of functions and distributions. Stability says that causal models should contain independencies of the former type, but not of the latter.

On the other hand, the invariance condition states that if causal variables are manipulated by interventions, then causal relationships among those variables are

Fig. 4.6 A causal structure. X causes Y and Z

not susceptible to change, i.e. the causal relation is invariant under a given class of interventions. *En passant*, what is the concept behind this 'invariant causal relation', i.e. *what* has to be invariant? The answer is: a detected *variation*. Thus formulated, invariance rules accidental or transient relations out. Invariance of the causal relation allows us to predict the value of Y, when we set different values for X, i.e. the same *variation* from X to Y is supposed to hold for different initial values of X.

To illustrate the different import of invariance and stability, let us look at an example. Consider Fig. 4.6 where the following equations hold (Pearl 2000: 49):

$$f_1 : Z = \alpha_1 X + \varepsilon_1 \tag{4.1}$$

$$f_2 : Y = \alpha_2 X + \varepsilon_2 \tag{4.2}$$

Z and Y will be independent, conditional on X, for all functions f_1 and f_2. However, suppose we modify the structure above by adding an arrow from Z to Y, to obtain Fig. 4.7. Suppose also that the following equations hold:

$$f_1 : Z = \gamma X + \varepsilon_1 \tag{4.3}$$

$$f_2 : Y = \alpha X + \beta Z + \varepsilon_2 \tag{4.4}$$

with $\alpha = -\beta\gamma$. Then, Y and X will be independent, because, by substituting $-\beta\gamma$ in f_2, Y does not depend on α any more. However, the independence between Y and X is not stable, because it disappears as soon as the equality $\alpha = -\beta\gamma$ is violated. So invariance requires invariability of structural coefficients (α, β, γ, etc.), whereas stability requires invariability of descriptive parameters, for instance, of the correlations among Z, Y, and X.

The moral is that if statistical and causal assumptions are satisfied, and if the model fits the data, then the conceptual hypothesis is accepted *in the specified background causal context*. This is what allows us to interpret variations causally. One issue needs to be underlined: causal assumptions are *constraints* on the variations. Invariance or stability tests play the role of constraints on the variation, that is they

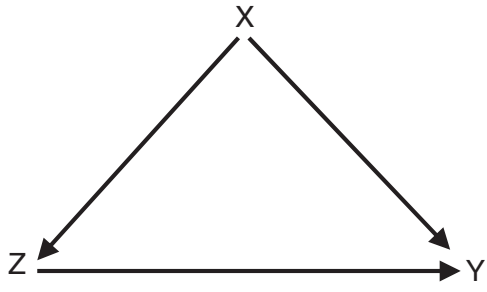

Fig. 4.7 A causal structure. X causes Z and Y, Z causes Y

are meant to ensure that the detected variation is causal rather than spurious or accidental. Let me emphasise once again that the testing stage is in fact about testing *variations*; in other words, the testing stage is not about detecting regular sequences of events, but about checking that detected variations satisfy certain conditions. Regardless of the preferred metaphysical theory of causality, when causal assumptions are satisfied, the causal claim stated in the conceptual hypothesis is confirmed (accepted, statistically speaking) and the relation between the *X*s and *Y* is confirmed to be causal. Statistical assumptions—as the difference between associational models and causal models shows—are not enough to guarantee the causal interpretation of variations.

If causal assumptions are met, then the structural model gives the causal structure. The causal structure stated in the conceptual hypothesis is disconfirmed in two cases: (i) when we cannot find a correspondence between the empirical data and the conceptual hypothesis, or (ii) when something goes wrong and one of the causal assumptions, for instance the invariance condition or no-confounding, fails. Indeed, this is what the hypothetico-deductive methodology of the vast majority of causal models in contemporary social science prescribes. Before illustrating by means of an example how all these assumptions work, let me add a methodological caveat. The claim that in case causal assumptions are met we get the causal structure dangerously hides a further assumption, namely that the data under analysis is correct.

One thing is to test the model and whether the conceptual hypothesis fits the data, and a different thing is to check the quality of data assuming the correctness of the model. In the first case, correctness of data is obviously assumed, except for outlying observations that the researcher might decide to exclude. The question of whether data is correct is nonetheless relevant but we cannot test at the same time the causal hypothesis *and* the correctness of data. In fact, to test whether data is correct we have to assume, in turn, the correctness of the model, or at least of the background knowledge the model is based upon. The correctness of the model then raises at least two further issues. (i) What does it mean for a model to be correct? (ii) Does the correctness of the model also imply its external validity? These are indeed highly valuable questions that would deserve a separate discussion. For the time being I will give the reader possible hints on how to deal with them. On the one hand, the philosophy of science has been debating about the concept of model for a long time. A model, according to some, represents reality. But this very same notion of representation turns out to be controversial. Those who feel uncomfortable with a strong notion of representation and do not want to commit themselves to the claim that a model enables us to attain the black box with laws of nature in it, can still adopt a weaker empiricist position according to which a correct model accurately describes the observable reality, *to the best of our knowledge*. On the other hand, social scientists are surely aware of the problem of extending the results of one study to a different population. Cook and Campbell (1979) gave very valuable warnings for practising researchers on the possible threats to external validity. Nonetheless, to be aware of the difficulties undermining the extension of result does not solve the problem at all. A step further can be made—and has to be made—in singling out

what theoretical conditions allow or prevent external validity. Let us now examine a case study to illustrate the various features of causal models.

A Case Study: Health and Wealth. In Chapter 1, I presented a study published in the *Journal of Econometrics*. By way of reminder, this study (Adams et al. 2003) aims at analysing possible causal paths between health and socio-economic status. The authors examine a sample from the population of elderly Americans aged 70 and older; the objective is to apply statistical methods (in particular, Granger-causality) to test for the *absence* of causal links from socio-economic status to health innovation and mortality, and from health conditions to innovation in wealth (see Fig. 4.8).

The graph above shows the general results of this study. The hypothesis of *no* causal link from health to wealth is accepted, as indicated by the thick black arrow; whereas the hypothesis of *no* causal link from wealth to health is rejected, as indicated by the dashed arrow. This case study is indeed particularly instructive because it exemplifies almost every feature sketched above and the hypothetico-deductive methodology employed in causal modelling.

(a) Knowledge of the causal context. Links between health and socio-economic status have been the object of numerous studies; this association holds for a variety of health variables and alternative measures of socio-economic status. Also, this association is already known to be valid across different populations. There has been considerable discussion about the causal mechanisms behind this association, but there have been relatively few experiments that permit causal paths to be definitively identified.

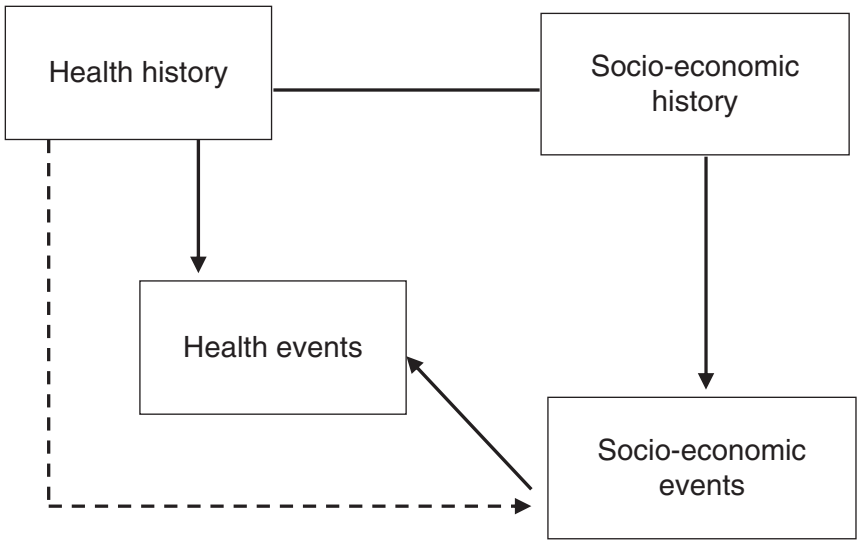

Fig. 4.8 Health and wealth relations

(b) Conceptual hypothesis. The authors expect the hypothesis of non-causality from health to wealth to be accepted, and the hypothesis of non-causality from wealth to health to be rejected. Such expectations come from contextual knowledge of the field and from results of similar studies. These hypotheses are put forward for empirical testing, as the hypothetico-deductive strategy discussed in section 3.2 advises.

(c) Extra-statistical assumptions: causal priority and causal ordering. According to Granger-causality, the past history of a variable Y_{t-1} causally determines the current value of the variable Y. So, to test whether health conditions have a causal impact on wealth presupposes the following assumption: health history determines wealth current events, and vice-versa when we test the causal relation from wealth to health. In fact, Granger (1969: 428) explicitly assumes that the future cannot cause the past. In this study, the asymmetry of causation plus determinism tell us that current values of the health variable can be exactly predicted from its history, which contains information about wealth conditions. Analogously, current values of the wealth variable can be predicted from its history, which contains information about health conditions.

(d) Statistical assumptions. The statistical analysis fits the approach of Granger (1969):

$$f(Y_t|Y_{t-1}) \equiv f_1(Y_{1t}|Y_{t-1}) \cdot f_2(Y_{2t}|Y_{1t} \cap Y_{t-1}) \cdot \ldots \cdot f_K(Y_{Kt}|Y_{K-1,t} \cap Y_{t-1}) \qquad (4.5)$$

Here Y_t denotes a K-vector of demographic, health and socio-economic variables for a household at time t, and Y_{t-1} is the information set containing the history of this vector through date t. Granger-causality is based on regression methods, hence it satisfies standard statistical assumptions.

(e) Causal assumptions. Structure of the causal relation and covariate sufficiency. For sufficiently brief time intervals, $f_K(Y_{Kt}|Y_{1t}, \ldots, Y_{K-1}, Y_{t-1})$ will not depend on contemporaneous variables, so 'instantaneous causality' is ruled out, and all the components in Y_t form a causal chain. Also, by focussing on first-order Markov processes, only the most recent history conveys information. Hence, all and only the components in Y_{t-1} are direct causes of Y_t.

(f) Invariance condition. In this study, the invariance condition is defined through the validity of the model: the model is valid for a given history Y_{t-1}, if $f(Y_t|Y_{t-1})$ is the true conditional distribution of Y_t given its history Y_{t-1}. That is, the conditional distribution must hold across the different panels analysed.

Thus, the causal interpretation of variations in health variables bringing about variations in wealth variables is tested by *imposing* the invariance condition as a condition for the model to be valid. More specifically, tests concern whether the hypothesised causal relation from health to wealth or from wealth to health invariably holds among different panels. Look at what happens when these tests are performed. To put it in a very simple way, there are two models to test. The first model describes *health variations given wealth variations*; the second model describes *wealth variations given health variations*. Invariance holds in the tests for the first model, whereas it fails in the tests for the second one.

In other words, Adams et al. were not expecting health status to have a significant impact on the income of retirees, and the empirical results are generally consistent with this expectation. To sum up, the failure of the invariance condition leads the authors to accept the conceptual hypothesis of non-causality from health to wealth. Although we do not reach the positive result of the presence of a causal relation, this claim is indeed instructive. Look at *where* causality is; causality is not at the output inferred from the rough data; instead, the causal structure stated in the conceptual hypothesis is put forward for empirical testing *within* the model, and it is accepted or rejected depending on whether consequences drawn from the conceptual hypothesis—the researchers' expectations—hold: in other words the conceptual hypothesis is accepted or rejected depending on the results of tests.

Moreover, look at the rationale of testing. The invariance condition is an expedient. It is a constraint that we impose on variations and that would allow us to interpret the model—e.g. health *variations* given wealth *variations*—causally. In other words, at the vary basis the rationale involves *measuring variations*, and the invariance exactly helps us in so doing.

A couple of remarks. First, as I said, the whole statistical set up is meant to confirm or disconfirm, i.e. to accept or reject, the conceptual hypothesis. The conceptual hypothesis *de facto* states a causal structure, which is accepted in a given causal context provided that statistical, extra-statistical, and causal assumptions are satisfied. Of course I am not maintaining that in *every* case study *all* these assumptions are made, but the list above provides a quite comprehensive inventory. As a matter of fact, causal assumptions are *constraints* on the causal relation, i.e. causal assumptions state what the characteristics of causal relations ought to be.

Second, much debate in philosophy of science is currently devoted to the Causal Markov Condition (CMC). Although I shall not discuss this issue in detail, I will try to pinpoint the general terms of the quarrel. This will enable me to stress the key trouble with the Causal Markov Condition. The debate mainly concerns whether or not CMC constitutes a necessary and sufficient condition for the causal interpretation of Bayesian nets; since in many cases CMC fails, some philosophers argue that CMC is not the decisive hypothesis (see for instance Hausman and Woodward (1999), Cartwright's reply (Cartwright 2002), and Hausman and Woodward's rejoinder to that (Hausman and Woodward 2004)).

Recall from section 3.1 the definition of a directed acyclic graph (DAG). A DAG is composed by a set of nodes and a set of arrows between them. Every node represents a variable, and arrows represent causal relations. I also mentioned that Bayesian nets often employ the terminology of kinship to denote various types of relations between the nodes in the graph. Simply put, CMC states that, in a directed acyclic graph, each variable is independent of all its non-descendants, conditional on its parents. To illustrate, consider Fig. 4.9 which represents dependencies among five variables (Pearl 2000: 15).

This graph describes the relationships between the season of the year (X_1), whether rain falls (X_2), whether the sprinkler is on (X_3), whether the pavement is wet (X_4), and whether the pavement is slippery (X_5). Consider now variable X_5. X_5 has only one parent, namely X_4; three ancestors, namely X_1, X_2, X_3; and no children.

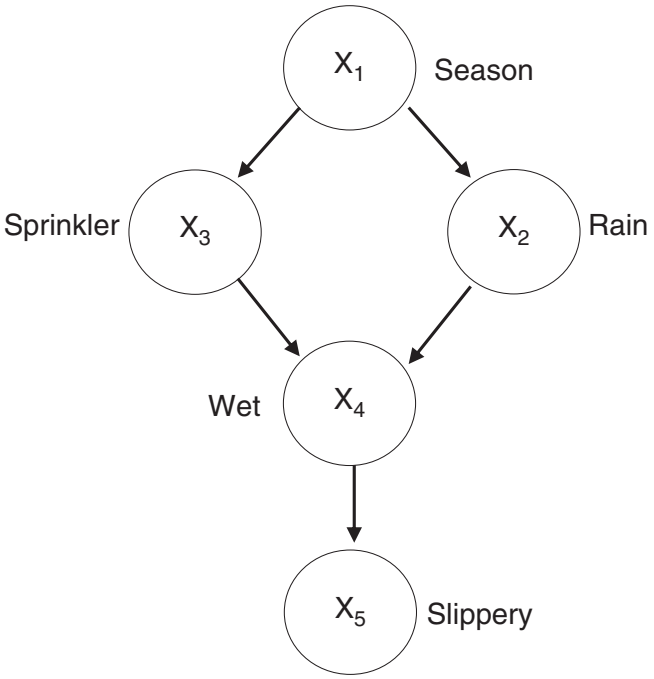

Fig. 4.9 A Bayesian net

CMC thus says that X_5 is independent of all its non-descendants X_1, X_2, and X_3, conditional on its parent X_4. In plain words, whether the pavement is slippery only depends on whether the pavement is wet. It does not matter whether this is due to the rain or to the sprinkler. In Bayesian nets, according to CMC, only parents are causes.

It is very important to bear in mind that the Causal Markov Condition is built into the definition of a Bayesian net. The point I want to make is that CMC *per se* cannot be what guarantees the causal interpretation, because a *non causal* interpretation can also be given to the Markov condition. In fact, although the Markov condition is built into the Bayesian net, the causal interpretation of a Bayesian net does not follow directly either. For instance, Bovens and Hartmann (2003) employ Bayesian nets in confirmation theory, in the evaluation of information in testimony, or in jury voting. In none of these applications is the causal interpretation involved, but Bayesian nets are nonetheless correctly and fruitfully used.

The key warning is that CMC is meant to guarantee causation under the assumption of the equivalence between causal dependence and probabilistic dependence. This assumption, however, is a debatable metaphysical assumption. In fact, the core of Bayesian nets methods, at least in Pearl's, Spirtes, Glymour and Scheines', as well as in Hausman and Woodward's approaches, is that Bayesian nets will bootstrap *causal* relations from mere *probabilistic* dependencies and independencies

among variables. In their manipulability theory, Hausman and Woodward (1999, 2004) offer a rather different argument for CMC. Hausman and Woodward's argument strongly relies on the concept of *modularity*. Modularity prescribes that effects should be manipulable by intervening *only* on direct causes (i.e. the parents), so that indirect factors (i.e. ancestors, or grandchildren) are left undisturbed. Cartwright (2000, 2002), however, vigorously argues against their defence. The foregoing discussion of the features of causal models has shown, instead, that several ingredients are needed in order to justify the correctness of causal interpretation. Besides causal assumptions, I also stressed the role of the causal context and of extra-statistical assumptions.

Let us now revert to the main track. Because I maintain that the causal interpretation crucially depends on the *causal* assumptions discussed above, the objection of vicious circularity is around the corner. My reply to this doubt is that, if there is a circle at all, it is virtuous and not vicious. To put causality in the assumptions in models is tantamount to saying two things: (i) that the evaluation of causal relations is relative to some conceptual framework, i.e. relative to available scientific theories and background knowledge, and (ii) that causal relations are confirmed thanks to specific formal constraints, i.e. statistical and causal assumptions.

The account of the guarantee of the causal interpretation given here has two apparent advantages. First, it avoids the metaphysical assumption according to which probabilistic dependencies are themselves causal. And second, it provides a rather complete list of the elements that grant the causal interpretation. It is worth noting that several authors mention the role played by the conceptual framework. For instance, the importance of the conceptual framework is also stressed by Humphreys (1989: §19, n. 33) or Kendall and Stuart.[10] As Haavelmo (1944: 22) puts it:

> We don't need to take the notions of cause and effect in any metaphysical sense. [...] The causal factors (or the 'independent variables') for one section of the economy may, themselves, be dependent variables in another section, while here the dependent variables from the first section enter as independent variables. [...] The notion of causal factor is of a relative character, rather then an absolute one.

Of course, this is not to deny the asymmetry of causation! But this simply shows that the notions of cause and effect depend on the conceptual framework. Fisher (1925: ch. 6) also makes a similar point when he says that:

> In no case, however, can we judge whether or not it is profitable to eliminate a certain variable unless we know or are willing to assume a *qualitative scheme* of causation. (My emphasis.)

Last, witness Suppes. (1970: 13):

> It is important to emphasize that the determination of a causal relationship between events or kind of events is always relative to a conceptual framework. There are at least three different kinds of conceptual frameworks within which it seems appropriate to make a causal claim. [...] One conceptual framework is that provided by a particular scientific theory; the

[10] Kendall, M. G., & Stuart, A. (1961). *The advanced theory of statistics*. Vol. 2, p. 317, quoted in Suppes (1970: 62), as well as in Humphreys (1989: §19, n.33).

second is the sort that arises in connection with a particular experiment or class of experiments; and the third is the most general framework expressing our beliefs with respect to all information available to us.

Unfortunately, their emphasis is not noticed enough in the literature, and consequently philosophers and statisticians engage in endless discussions *only* about technicalities of formal constraints, which very often presuppose unjustified metaphysical assumptions. It seems to me that the references just mentioned are very good evidence in support of the emphasis I bestow on background knowledge: the conceptual hypothesis comes from background theories and knowledge, and the qualitative scheme of causation—as Fisher calls it—thereby described is put forward for testing. The causal interpretation crucially depends on formal constraints but, as just mentioned, the role of background knowledge is not negligible either. Indeed, this is something Durkheim (1897) had already taught us—that no method will allow us to obtain mechanically a causal relation, unless it is accompanied by an elaboration of the mind, namely by a conceptual framework. The references above are far from being an appeal to authority. Rather, they are meant to stress the nonreductivist character of causal modelling. We cannot reduce causation to statistics and this is why particular attention has to be paid to background knowledge and to the formulation of the conceptual hypothesis in modelling.

Before closing this chapter, I wish to draw the reader's attention to another tricky methodological point. As I just mentioned, Bayesian nets consider as causes only parents, because according to the Causal Markov Condition children are independent of every ancestor, conditional on their parents. This seems plausible in Pearl's example mentioned above, where, in the ultimate analysis, we would not consider

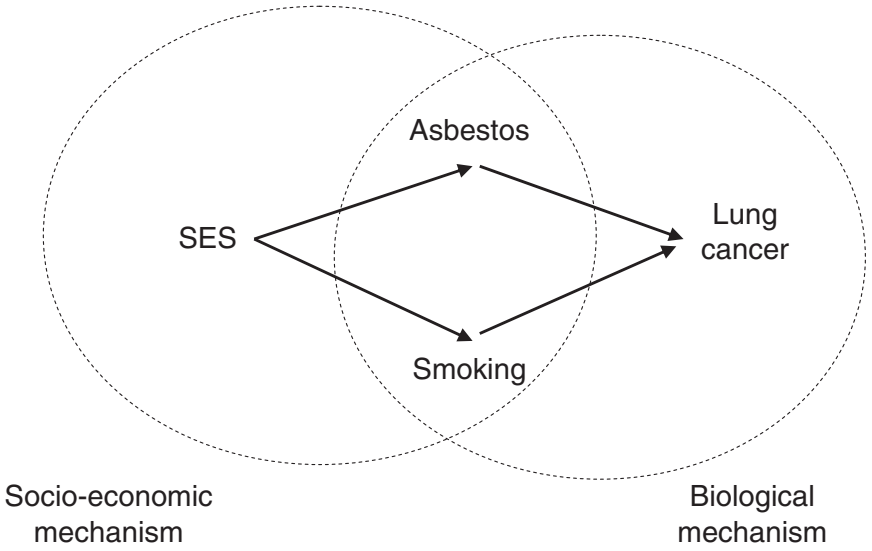

Fig. 4.10 Relations between socio-economic status, asbestos, smoking, and lung cancer

the season as a cause of the slippery surface. But consider a different DAG, for instance one in which cancer has two causes, say smoking and asbestos, which in turn have a common parent, say socio-economic status (see Fig. 4.10).

As Russo et al. (2006) discuss in some detail, this graph depicts two different causal mechanisms—a biological mechanism and socio-economic one. An epidemiologist would probably be interested in the biological mechanism, i.e. in the relations holding between smoking, asbestos exposure and cancer rate. However, a demographer might be further interested in understanding what part of the population has higher cancer rates, conditional on other characteristics, such as smoking behaviour or asbestos exposure. Thus, the model needs to be broadened in order to consider socio-economic status as an (indirect) cause of cancer. However, in the Bayesian nets approach, the socio-economic status would not be a cause (even indirectly) of cancer because it is an ancestor and not a parent. This problem is not just a linguistic quarrel—it goes far beyond the terminological problem of referring only to parents as 'causes'. This last example raises several issues. Do, can, or should ancestors count as causes? If so, why and in what cases? Do different types of variables (social, demographic, biological, etc.) equally count as causes in a DAG? How do we make it plausible to mix different types of variables in the same model? I shall answer those questions in some detail in Chapter 6, where we will deal with modelling causal mechanisms.

Chapter 5
Methodological Consequences: Objective Bayesianism

Abstract This chapter argues that objective Bayesianism is the interpretation of probability that best fits causal modelling in the social sciences. An overview of the leading interpretation is offered. The argument for objective Bayesianism is structured in three steps: first, I maintain that we need an interpretation that can account for single-case and generic probabilistic causal claims; second, I show that the empirically-based and objective Bayesian interpretations are apt to do this; third, I claim that the objective Bayesian version is preferable on the grounds that it leaves no room to arbitrariness, enlightens the design and interpretation of tests, and is a better guide for action.

Keywords Interpretations of probability; classical and logical interpretation; physical interpretation; Bayesian interpretations: subjective, objective, empirically-based; single-case causal claims; generic causal claims.

Introduction

As shown in Chapter 2, *probabilistic* theories of causality have been proposed over the last few decades; I discussed the limits of these approaches and suggested considering *causal modelling* instead. As a matter of fact, causal models presented in Chapter 3 are *probabilistic* in character too, in the sense that they make use of probabilities in an essential manner. Causality is closely connected with the *interpretation* of probability and this chapter investigates what interpretation best fits causal modelling.

In section 5.1, I present two different types of probabilistic causal inference that causal modelling handles. Notably, I introduce the distinction between 'generic' and 'single-case', that will be thoroughly dealt with later in Chapter 6. Whilst generic casual claims pertain to the population, single-case causal claims refer to particular individuals. Yet, they are both probabilistic and call for an appropriate interpretation of probability. Section 5.2 offers an overview of the leading interpretations of

probability. The goal is not to furnish a complete and exhaustive presentation, but rather to give the reader enough background to understand the choice of a *Bayesian* interpretation of probability.

Section 5.3 presents an argument to support Bayesianism, and, particularly, the empirically-based and objective versions, where epistemic probabilities are 'frequency-driven'. Two points above all will be underlined: firstly, Bayesianism is a suitable interpretation for different categories of causal claims, and secondly, Bayesianism allows 'learning from experience'. Finally, section 5.4 goes a step further and defends the objective Bayesian interpretation over the empirically-based version on the following grounds: objective Bayesianism performs better (i) in the design and interpretation of tests and (ii) in guiding action.

5.1 Probabilistic Causal Inferences

As we have seen in previous chapters, social science research is interested in the various aspects and characteristics of populations. We have also seen that quantitative social research is made up of a plurality of methods and approaches. Yet, in their common objective of understanding, predicting and intervening on society, *causal* relations play a major role.

On the one hand, the examples discussed earlier in sections 2.1 and 2.2 already pointed to the fact that causal relations may refer to quite different things: the tendency of an event to cause another, the actual causal effectiveness of a given event, the frequency of occurrence of a causal relation, etc. On the other hand, although causal models presented in Chapter 3 are primarily concerned with generic causal claims, such generic causal relations have to be applicable to the single case too. In Chapter 6, I will discuss the generic vs. single-case distinction thoroughly. For the time being, an example involving a physiological mechanism will suffice to grasp the idea. Consider epidemiological studies about the effects of tobacco consumption on lung cancer. For policy reasons (among others), we are interested in determining the incidence of tobacco consumption on lung cancer, namely we are interested in a *generic causal relation* that holds at the population level. But we might also be interested in a *single-case causal relation*—for instance, in Harry's chance of developing lung cancer, given that he smokes, or in the probability that smoking actually caused him to develop cancer.

The issue of the interpretation of probability becomes particularly interesting because causal claims at both levels are probabilistic, yet they express quite different things. Whilst generic causal claims seem to be primarily concerned with frequency of occurrence, single-case causal claims seem to express a (rational) belief about what did or will happen. Causality thus appears to be closely connected to the interpretation of probability, because, if probability has to be applied, it also has to be interpreted. Therefore, it is crucial to ask: to what extent is the interpretation of probability sensitive to those shifts of meaning? What interpretation should we adopt in order to account for both generic and single-case causal claims? In this chapter we will see which interpretation, among the available ones, is apt to do this job.

Causality, as we have seen so far, is a highly problematic notion, but so is probability. In fact, despite the substantial core of agreement on the syntax of probability language, namely on Kolmogorov's axiomatization, philosophers still vividly debate on its semantics, namely on the *interpretation* of probability.

5.2 Interpretations of Probability

Early attempts to give a rigorous mathematical account of the notion of probability trace back to the French mathematicians Blaise Pascal and Pierre Fermat in the seventeenth century. However, probability theory was much later axiomatised by Kolmogorov in 1933. Simply and informally, the axioms state that (i) probabilities are non-negative real numbers, (ii) every tautology is assigned value 1, and (iii) the sum of the probabilities of two mutually inconsistent sentences is equal to the probability of their disjunction. The conditional probability of A given B is written $P(A|B)$ and is defined as $P(A|B) = P(A\&B)/P(B)$, with $P(B) \neq 0$. Bayes' theorem follows from the axioms and from the definition of conditional probability. The theorem governs the inversion of a conditional probability and relates the posterior probability of B given A to the probability of A given B, provided that the prior probabilities of A and B are known or that a conventional procedure to determine them is accepted. In symbols, Bayes' theorem states that $P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$, with $P(A) > 0$. For a lucid exposition also accessible to a nonspecialist audience, see, for instance, Howson and Urbach (1993) and Suppes (2002: ch. 5).

This formal apparatus allows us to calculate probabilities. But how is probability to be interpreted? What is its meaning? Thousands of pages have been written on the interpretation of probability. So many interpretations have been proposed that to give a complete and exhaustive account of all of them is not just a challenging but rather an impossible enterprise, especially in a few pages. Very good introductions are those of Cohen (1989), Gillies (2000), Suppes (2002), Hájek (2003), to which I direct the reader. In the following, I limit myself to providing a very brief overview of the major interpretations; my aim is to give the reader enough background to understand the choice of a Bayesian interpretation, and in particular of a Bayesian interpretation in which degrees of belief are shaped upon available evidence. Here are the leading contenders:

(i) Classical and logical interpretation.
(ii) Physical interpretation: frequentist and propensity.
(iii) Bayesian interpretations: subjective interpretation, objective and empirically-based interpretation.

5.2.1 Classical and Logical Interpretation

Briefly put, in the *classical interpretation* probability is defined as the ratio between the number of favourable cases and the number of all equipossible cases. The easiest

way to grasp the meaning of favourable and equipossible cases is to think of dicing. The six sides of a dice constitute the probability space—i.e. the six possible outcomes. Assuming that the dice is not biased, the six sides are all equipossible—i.e. they all have the same probability to come out. The favourable case is then the event we are interested in. For instance, the probability that an even number will come out is given by the number of favourable cases (sides 2, 4 and 6) over the total number of the equipossible cases (for an unbiased dice: 1, 2, 3, 4, 5 and 6) which gives 3/6, i.e. 1/2.

Firstly theorized by Laplace (1814), classical interpretations are also those of Pascal and Jakob Bernoulli. Probability values are assigned in absence of any evidence—this is why the probability of an event is simply the fraction of the total number of possibilities in which the event occurs. The notion of equipossibility is expressed by the principle of indifference, a coinage due to Keynes. The principle states that whenever there is no evidence favouring one possibility over another, these possibilities have the same probability. The classical interpretation seems especially well suited to games of chance, although it suffers the problem of circularity, for equipossible means equiprobable, hence 'probable' is not properly explicated. A second traditional objection is that the classical interpretation is of scarce applicability in science; in fact, adopting the classical interpretation we have no meaningful way of expressing knowledge of the population probabilistically, nor of evaluating individual hypotheses.

A generalisation of the classical interpretation is the so-called *logical interpretation*, advanced by Keynes (1921), Jeffreys (1939), and mainly by Carnap (1951). This interpretation rests on the idea that probabilities can be determined a priori by an examination of the space of possibilities. However, it allows for cases in which possibilities are assigned unequal weights, and probabilities can be computed whatever the available evidence may be. In fact, the main aim of the logical interpretation is to provide an account, as general as possible, of the degree of support or confirmation that a piece of evidence e confers upon a given hypothesis h. According to Carnap's notation, the c-function $c(h|e) = q$ precisely expresses this idea. Logical interpretations—and in particular Carnap's—calculate subjective probability values upon knowledge of objective probability values, so these interpretations might also be classified next to empirically-based or objective Bayesian approaches, which I will shortly introduce.

5.2.2 Physical Interpretation: Frequentist and Propensity

According to the *physical interpretation*, probability values are quantitative expressions of some objective feature of the world.[1] Within the physical interpretation, the frequency and the propensity interpretations are typically distinguished. A simple

[1] Traditionally, the literature classifies the frequency and propensity interpretations as 'objective' interpretations. Here, following Williamson (2005a), I prefer the label 'physical' for reasons explained below.

version of *frequentism*, due to Venn (1876), states that the probability of an attribute A in a finite reference class B is the relative frequency of the actual occurrence of A within B. Further developments of frequentism are due to Reichenbach (1949) and von Mises (1957), who consider infinite reference classes and identify probabilities with the limiting relative frequencies of events or attributes therein. This second sort of frequentism is also advocated by Salmon (1967). Limiting relative frequencies serve, in his approach, in determining the probability of evidence. Frequentism mainly suffers the problems of assigning a probability value in the single case because to compute it we have to pick out the right actual or hypothetical reference class.

The *propensity interpretation* also is located in the 'physical realm', since probability is 'in the world', so to speak, rather than 'in our heads' (as in subjectivist approaches), or in logical abstractions (as in the classical and logical interpretation). Probability is here conceived of as a physical propensity, or disposition, or tendency of a given type of physical situation to yield an outcome of a certain kind, or to yield a long run relative frequency of such an outcome. The propensity interpretation was advanced by Popper (1957, 1959), who was motivated by the desire to make sense of single-case probabilities, for instance in quantum mechanics. A typical objection to propensity theories is that statements about single-case propensities are untestable, therefore they are metaphysical rather than scientific.

In sum, physical interpretations will not suit our needs either because they do not make sense of the single case or because they are of scarce applicability in the evaluation of individual hypotheses. In fact, the probability of a single and unrepeatable event has no meaning for a frequentist who instead evaluates the probability of an event that may occur indefinitely in identical conditions. For the same reason, within the physical framework, we cannot evaluate the probability that a given hypothesis will materialise, but only the probability of observing the sample if the hypothesis is true (on this point see also Courgeau (2004b) and later section 5.4).

5.2.3 Bayesian Interpretations

Before presenting in more detail the three versions of the Bayesian interpretation, it is worth spending a few words on Bayesianism as an epistemological stance. There are two main assumptions behind Bayesianism. First, scientific reasoning is reasoning in accordance with the formal principles of probability theory, and second, Bayesianism provides an account of how we do or should learn from experience. The formal apparatus of probability theory serves to impose coherence constraints on rational degrees of belief and uses conditionalisation as a fundamental probabilistic inference rule for updating probability values according to Bayes' theorem. Thus, Bayesianism allows inductive reasoning from data: it purports to explain probabilities of hypotheses in the light of data. Bayesianism, as an epistemological position about scientific reasoning, is accompanied by an interpretation that takes probabilities to be rational degrees of belief. Such an interpretation comes in two main

versions: (i) subjective and (ii) objective and empirically-based. Unlike physical interpretations, Bayesian interpretations take probabilities to be quantitative expressions of an agent's opinion, or degree of belief, or epistemic attitude, or anything equivalent. Thus, in Williamson's (2005a) terminology, they are mental rather than physical, and for this reason they are sometimes referred to also as 'epistemic' probabilities. Bayesian interpretations will then differ in the constraints they impose on rational degrees of belief, that is, in the freedom or arbitrariness a rational agent is allowed in setting those probability values.

Subjective Bayesianism. First advances of the *subjective* interpretation are due to Ramsey (1931) and de Finetti (1937). De Finetti's viewpoint is paradigmatic of personalistic approaches, for he firmly states that probability does not (physically) exist, and that it is possible to reconstruct and deduce probability theory just relying on the subjectivist interpretation (de Finetti 1993: 248 ff.). In subjectivist approaches probabilities are typically analysed in terms of betting behaviour: probabilities are identified with the announcement of the betting odds that a rational agent is willing to accept.

A *Dutch book* (against an agent) is a series of bets, each acceptable to the agent, but which collectively guarantee her loss, whatever happens. Two Dutch book theorems then follow: (i) If an agent's subjective probabilities violate the probability calculus, then she is liable to a Dutch book; and, conversely, (ii) if an agent's subjective probabilities conform to the probability calculus, then no Dutch book can be made against her. An agent is then called irrational when a Dutch book is performed against her. That is to say, obedience to the probability calculus is a necessary and sufficient condition for rationality. It is typically objected that subjectivist accounts lead to arbitrariness, that is, they are *too* subjective. In fact, two agents may assign different probability values to the same event and be equally rational, provided that they do not violate the probability calculus. It is worth pointing out that de Finetti's betting interpretation derives probabilities from utilities and rational preferences. The intimate link between utility of outcomes and probabilities is even more remarkable in the approaches of Savage (1954) and Jeffrey (1966). The main idea of the utility interpretation is that probabilities and utilities can be derived from preferences among options that are constrained by certain putative consistency principles. Unfortunately, subjectivist approaches do not seem a viable solution for our purposes because we lack a meaningful interpretation for probabilities at the generic level, although they are well suited to express probability in the single case.

Objective and empirically-based Bayesianism. A solution to the objection of arbitrariness is attempted in the *objective* and *empirically-based Bayesian* interpretations. These versions of Bayesianism require that further constraints beyond coherence be satisfied before an agent's degrees of belief can be considered rational. An early proponent of objective Bayesianism is Jaynes (1957, 2003). Two types of constraints may be distinguished: empirical and logical. Information and lack of information about the world, respectively, should be taken into account in shaping epistemic probabilities. Jaynes puts forward a maximum entropy principle, which may be thought of as an extension of the principle of indifference. Salmon (1967)

emphasises the role of empirical constraints and requires knowledge of relative frequencies in order to assign prior probabilities, therefore his account can be put next to empirically-based Bayesianism. Recently, Williamson (2005a) attempts a novel defence of objective Bayesianism, requiring that *both* empirical and logical constraints be satisfied. Possible objections concern the use of frequencies and the applicability of a maximum entropy function.

There is another difference between the empirically-based and the objective interpretation. Whilst in the empirically-based interpretation knowledge of frequencies is sufficient to determine a degree of belief, the objective interpretation also imposes logical constraints. Logical constraints become particularly relevant when evidence is not enough to determine a unique degree of belief. In these cases, in fact, objective Bayesianism prescribes choosing the most equivocal or middling probability value, whereas under the empirically-based interpretation the agent is free to choose any value.

The objective and empirically-based Bayesian interpretations seem to fit our purposes much better than other interpretations because they enable a twofold interpretation of probability. Single-case probabilities are interpreted as rational degrees of belief constrained, among other things, by knowledge of frequencies. The frequency interpretation thus turns out to be compatible, and in a way complementary, to the empirically-based and to the objective interpretations and thus offers an appropriate meaning in generic causal claims that instead state frequency of occurrence. In the following, I shall defend those positions thoroughly.

5.3 The Case for Frequency-Driven Epistemic Probabilities

After this brief overview, let me recall the question to be addressed in this section. What interpretation of probability best fits different sorts of probabilistic causal claims? In a nutshell, the answer I attempt is the following: *Bayesianism*, and particularly empirically-based or objective Bayesianism, are the interpretations that best fit causal modelling in the social sciences. In this section, I will argue more thoroughly in favour of the empirically-based and objective Bayesian interpretations.

Before running the argument, it is worth pointing out that the purpose is not to refine any of the existing versions of Bayesianism. Rather, the intention is to explain the motivation for, and the soundness of, the Bayesian approach. As I said above, causality forces us to take a position *vis-à-vis* the interpretation of probability: if probability has to be applied, it has also to be interpreted. However, the choice of the *interpretation* is more concerned with the discussion of foundational issues than with technicalities. Consequently, I will focus on the *philosophical* reasons why Bayesianism is a sensible interpretation for the epistemology I advanced in Chapter 4 and for the practice of causal modelling. Notably, I shall mainly address two foundational issues: the soundness of a twofold interpretation of probability, and the role of experience in Bayesian reasoning.

5.3.1 Probability as a Twofold Concept

The concept of probability is essentially a dual concept that endorses an objective and a subjective side. This is not tantamount to saying that a principle of tolerance should allow all interpretations equally. Rather, this claim states that the two sides serve different purposes.

The duality of the probability concept is analysed at length in Hacking's *The emergence of probability*. Hacking maintains that ever since the first formulations of the probability theory, at the time of the famous epistolary exchange between Pascal and Fermat, 'probability' was meant as degree of belief *and* as the tendency of a chance device to display stable relative frequencies. Hacking's historical thesis is challenged, however. According to Gillies (2000: 18) the emergence of duality appeared a bit later. It surely started with Laplace, and according to Daston (quoted in Gillies 2000: 10) duality traces back to Poisson, Cournot, Ellis. Whatever the correct tenet—Hacking's or Gillies' historical claim—it is a matter of fact that the duality of the concept of probability has a long history and that, moreover, several recent accounts employ both the subjective and objective concept of probability.

So, the 'Janus-faced' aspect of probability, as Hacking and Gillies call it, appears to be tenable. Probability is called 'Janus-faced' because, as the Roman god has two faces—one looking at the past and the other at the future, so probability has two sides—subjective and objective. That is to say, a twofold view is defensible and different interpretations may better fit different contexts. Gillies himself is a defender of pluralism. He reinforces the pluralist view of probability by arguing that there are two broad areas of intellectual study which require different interpretations of probability. An *epistemological* notion of probability (that is, in the jargon here adopted, rational degrees of belief) is appropriate for the social sciences, whereas an *objective* (that is, in the jargon here adopted, physical) notion is appropriate for the natural sciences (Gillies 2000: 187 ff.). Notice, however, that the kind of pluralism here advocated differs from Gillies' in that it requires two distinct concepts of probability in the same domain.

To show how 'Janus-faced' probability is involved in many contemporary interpretations, I will borrow Salmon's (1988) distinction between *frequency-driven* (F-D) accounts of subjective probability and *credence-driven* (C-D) accounts of objective probability. Let me spell out the meaning of F-D and C-D accounts first; I will then go through some approaches in order to highlight the fact that they all share two characteristics: (i) they all employ a twofold concept of probability; and (ii) to shape epistemic probabilities they all resort to frequencies.

The difference between F-D approaches and C-D approaches can be stated as follows. Given that there are two kinds of probabilities—says Salmon (1988: 15 ff.)—the question to be addressed is how they relate to each other, that is, we have to understand what is the relationship between epistemic probabilities and frequencies. In F-D accounts, frequencies play a major role in determining subjective probabilities, whereas C-D accounts instead go the other way round: physical probability is based on belief. To borrow Salmon's words, the whole point is to make clear 'who is in the driver's seat': physical or subjective probability?

Salmon discusses Ramsey's approach as paradigmatic of the F-D accounts, and Mellor's and Lewis' as paradigmatic of the C-D accounts. To act above board, Salmon obviously sympathises with F-D accounts, as indeed I do. Yet, we are not the only two partisans. Many subjectivists (Ramsey, Carnap, Salmon, van Fraassen) of different sorts ultimately rely on frequencies in order to shape epistemic probabilities. Lewis' Principal Principle will instead be instructive in clarifying the nonarbitrariness of the empirically-based and objective Bayesian approaches.

5.3.2 Frequency-Driven Epistemic Probabilities

Ramsey (1931). In "Truth and probability", Ramsey's intent is to provide an understanding of the concept of degree of belief and an account of the logic of partial belief. In the Foreword, Ramsey expresses his doubts about whether a single interpretation of probability be appropriate in different contexts such as logic and physics. Let us anticipate his solution and then see how he reached it.

There are two viable interpretations—an interpretation in terms of class-ratio, that is the frequency interpretation, and an interpretation as a calculus of consistent partial belief, that is the subjective Bayesian interpretation. On the one hand, we expect that there be some intimate connection between the two, and on the other, there must be some explanation of the possibility of applying the same mathematical calculus to both (Ramsey 1931: 25ff.).

There are two types of connections between (partial) belief and frequencies. First, experienced frequencies often lead to corresponding (partial) beliefs; and second, (partial) beliefs lead to the expectation of corresponding frequencies. However, (partial) belief cannot be uniquely connected with any actual frequency, but rather with a hypothetical or ideal frequency. The first type of connection is what interests us most. Ramsey (1931: 26) says:

> It is this connection between partial belief and frequency which enables us to use the calculus of frequencies as a calculus of consistent partial belief. And in a sense we may say that the two interpretations are the objective and subjective aspects of the same inner meaning, just as formal logic can be interpreted objectively as a body of tautology and subjectively as the laws of consistent thought.

Belief, in other words, is shaped upon frequencies, whether actual or hypothetical. Let us now see what kind of reasoning justifies the solution.

Ramsey concedes that modern science uses probability theory in the sense of the frequency theory, that is a ratio between the numbers of two classes or the limit of such a ratio. However, probability can also be given a different meaning, that is as of partial belief. This is the case, according to Ramsey (1931: 8), both in ordinary language and in the thought of "other great thinkers" (ibidem). Ramsey does not develop more on this point, unfortunately. But it seems plausible that what he had in mind are cases in which we express a belief about what will or did happen—for instance, the belief that Harry will develop lung cancer given that he smokes or that he developed cancer because of his smoking—rather than stating a proportion between

two classes—for instance, between the number of lung cancers within the class of smokers. Ramsey's working hypothesis is then the following (Ramsey 1931: 8):

> It may be that, as some supporters of the frequency theory have maintained, the logic of partial belief will be found in the end to be merely the study of frequencies, either because partial belief is definable as, or by reference to, some sort of frequency, or because it can only be the subject of logical treatment when it is grounded on experienced frequencies.

The answer will come through an investigation of the logic of partial belief. In order to do develop a logic of partial belief, we need a notion, at least approximate, of partial belief and a way to measure it. Belief, as a psychological variable, has obvious measurement difficulties. But, according to Ramsey (1931: 15), there is indeed a way to measure belief, that is considering beliefs "qua basis of action". This is possible if we consider beliefs to be *probabilities*. In particular, this way of measuring belief can be explicated in terms of betting behaviour, which is quite a standard strategy among subjective Bayesians. Ramsey (1931: 22) then shows that his account of the logic of partial belief leads to laws of probability as laws of consistency, that is, in Bayesian terms, obedience to the axioms of probability calculus is a necessary (and sufficient, for a subjective Bayesian) condition for rationality.

The circle has now to be closed. If, as Ramsey has argued and shown, the calculus of probability can be interpreted consistently in two different ways, then there must be a connection between the two: the calculus of frequencies can be used as a calculus of consistent partial belief.

Carnap (1951). In *The logical foundations of probability* Carnap distinguishes two concepts of probability (§9, §10A and §10B). *Probability*$_1$ denotes the weight of evidence or the degree of confirmation, and *probability*$_2$ denotes relative frequency. Probability$_1$ can be explicated in three ways: (i) as a measure of evidential support given to a proposition h in the light of a different proposition e (§41A); (ii) as a fair betting quotient (§41B); and (iii) as an estimate of relative frequency (§41C and §41D). Carnap's strategy is to explicate the concept of probability$_1$ as an estimate of relative frequency (iii) *via* the concept of a betting quotient (ii), in turn explicated *via* the measure of evidential support (i). For instance, suppose that the relative frequency of the attribute M in a class to which b belongs to is known to be r, then the fair betting quotient for the hypothesis that b is M, and hence the probability$_1$, of this hypothesis is r.

In the opening of §8, Carnap makes it clear that the concepts of confirmation he will deal with are semantic and logical. In particular, the quantitative concept of confirmation, i.e. *probability*$_1$, has two arguments—the hypothesis and the evidence—although the latter is often omitted, and in §10A Carnap stresses the importance of the evidential component of probability$_1$.

Consider now Carnap's c-function. The result of an inductive inference has the structure of the c-function: $c(h|e) = q$, where propositions h and e are the hypothesis and the evidence, and q is a real number in the interval [0, 1]. q obeys the axioms of probability theory, i.e. q is a probability value; q here represents the *degree of confirmation* in the hypothesis h on the basis of the *evidence* e (see §8 and §55). It is important to bear in mind that the evidence e represents available experimental

or observational evidence, viz. what we *know* about the world and on the basis of which probability values in inductive inferences are shaped. Now, if e is also a probability statement, then probability has to be interpreted as probability$_2$, that it is to say, probability$_1$ is shaped on frequencies in e. Carnap reiterates these ideas in §§ 41–42, and in §42B says:

> Many writers since the classical period have said of certain probability statements that they are 'based on frequencies' or 'derived from frequencies'. Nevertheless, these statements often, and practically always, if made before the time of Venn, speak of probability$_1$, not of probability$_2$. In our terminology they are probability$_1$ statements referring to an evidence involving frequencies. [...] in these cases the probability determined with the help of a given frequency and its value is either equal to or that of the frequency.

Salmon (1967). In *The foundations of scientific inference*, Salmon tackles the old problem of induction. Induction was famously criticised by Hume, who was seeking to understand how we acquire *knowledge* of the unobserved. Salmon then draws a distinction between knowledge and belief: knowledge, unlike mere belief, is founded upon evidence; that is, we need to provide a rational justification for it. Salmon is here raising a *logical* problem: the problem of understanding the *logical* relationship between evidence and conclusion in correct inductive inferences. As is well known, inductive inferences cannot establish true conclusions but only *probable* conclusions from true premises. Because we are dealing with *probability* statements, we also have to provide an interpretation for them. Salmon (1967: 48–50) considers two basic meanings of the concept of probability: as frequency and as rational degree of belief.

The frequentist concept, he argues, is not a suitable interpretation of probabilistic results in inductive inferences. The reason is this. Given that under the frequentist interpretation "the probable is that which happens often and the improbable is that which happens seldom" (Salmon 1967: 48), if we claim that inductive inferences are probable in this sense, we would be claiming that inductive inferences with true premises often have true conclusions, although not always.

However, Hume's critique of induction has proved two things: (i) inductive inferences cannot establish their conclusion as true even if the premises are true; and (ii) inductive inferences cannot establish conclusions as probable in the frequentist sense. Instead, the concept of probability as degree of belief is more promising. In particular, *rational* degrees of belief work better in making clear the meaning of 'probable' in conclusions of inductive inferences: to say that a statement is probable means that one would be justified to believe it and that the statement is supported by evidence. Moreover, in inductive inferences, rational degrees of belief are *objectively determined* by evidence. Evidence supports a statement depending on the inductive rules we adopt; according to Salmon, induction by enumeration is the basic inductive rule for this purpose, and allows us to infer the limit of the probability value from the virtually infinite sequence of possible outcomes, i.e. the limit of the relative frequency (Salmon 1967: 96–98). The problem now is how we use this evidence in inductive inference to confirm hypotheses.

The solution comes from the probability calculus itself: Bayes' theorem. Bayes' theorem, says Salmon (1967: 117), "provides the appropriate logical schema to

characterise inferences designed to establish scientific hypotheses" (ibidem). However, Bayes' theorem poses difficulties of interpretation. The formal scheme of Bayes' theorem requires prior probabilities—but what are those probabilities? Salmon's answer is: frequencies. "[…] The frequency interpretation of probability can be used to approach the prior probabilities of scientific hypotheses". In sum, rational degrees of belief are based upon the frequencies

van Fraassen (1983). In his article on the justification of subjective probabilities,[2] van Fraassen distinguishes two uses of the concept of probability. The first refers to the frequency interpretation: probability statements are about actual frequencies of occurrence of events. The second refers to the subjective interpretation and serves to formulate and express our opinion and the extent of our ignorance concerning matters of fact. According to van Fraassen, any satisfactory account of probability should explicate both uses as well. However, whilst proponents of Bayesian views have done quite well, frequentists have not. On the one hand—he says—Bayesians have successfully shown that obedience to the probability calculus provides a necessary criterion of rationality. On the other, frequentists have never succeeded in meeting major critiques of their failure to account for the subjective use of probability. As van Fraassen (1983: 295) puts it, he will attempt to redress the balance and to do so he will try to demonstrate that the observance of the probability calculus in expressions of opinion—i.e. rational degrees of belief—is equivalent to the satisfaction of a basic *frequentist* criterion of rationality: potential calibration.

According to van Fraassen rational degrees of belief, once expressed, are evaluated in two ways. One question is whether they are *reasonable* and the other is whether they are *vindicated*. From a Bayesian standpoint—recall—subjective probabilities are equated with the betting odds the agent is willing to accept. Thus vindication consists in gaining, or at least not losing no matter what happens, as a consequence of such bets. Dutch book theorems then state that vindication will be a priori precluded if, and only if, probability values do not satisfy the probability calculus. This way coherence is a minimal criterion of reasonableness connected with vindication, in particular, the possibility of vindication is taken as a requirement of reasonableness. However, we can adopt a different strategy to evaluate reasonableness and vindication of degrees of belief: explicate vindication in terms of calibration, and the possibility of vindication in terms of potential calibration.

Simply put, calibration describes the behaviour of a forecaster. A good forecaster should be informative—i.e. probability values assigned to a statement should approach 0 or 1—and well calibrated, where calibration is a measure of how reliable the forecaster is: the higher the frequency of true predictions, the more reliable the forecaster. A forecaster, then, will be perfectly calibrated when she chooses the correct reference class and estimates frequencies that happen to be correct. Potential calibration, or frequency coherence, concerns the extension of the set of propositions beyond the initial set to which we attached probability values and that turned out to be calibrated. The idea of being potentially calibrated is that it is

[2] In this particular paper van Fraassen uses a slight different terminology. He oft uses 'personalistic' for 'subjectivist', and 'attitudes' or 'opinions' for 'rational degrees of belief'.

possible to vindicate rational degrees of belief a priori. This is possible, according to van Fraassen, because obedience to the probability calculus is equivalent to this frequency-coherence criterion. Therefore, a forecaster can safely shape rational degrees of belief on frequencies insofar as she stays potentially calibrated. In other words, van Fraassen is claiming that degrees of belief can be given a frequentist vindication, and in this way he binds subjective probabilities to frequencies..

5.3.3 Credence-Driven Physical Probabilities

Lewis' Principal Principle (1971). Lewis' *Principal Principle* is not an F-D account but rather a C-D one. Thus the Principle (Lewis 1971: 266):

> Let C be any reasonable initial credence function. Let t be any time. Let x be any real number in the unit interval. Let X be the proposition that the chance, at time t, of A's holding equals x. Let E be any proposition compatible with X that is admissible at time t. Then, $C(A|XE) = x$.

For instance, if A is the proposition that a coin tossed at time t will fall heads, X is the proposition that the chance of A at time t is x, and E is the available evidence that does not contradict X; the Principal Principle then says that x equals the actual degree of belief that the coin falls heads, conditionally on the proposition that its chance of falling heads is x. In other words, the chance of A equals the degree to which an agent believes in A.

The difference between F-D accounts and the Principal Principle is subtle but fundamental. According to F-D accounts I examined, knowledge of objective probabilities is used to determine a rational degree of belief; that is to say, subjective probabilities must ultimately be based upon physical probabilities, in particular, upon knowledge of observed frequencies. On the other hand, in the Principal Principle the chance of an event is the *chance* of the truth of the proposition that holds in just those worlds where *this* event occurs. It is apparent that Lewis' concept of objective chance differs widely from Popperian propensities in that it is time-dependent and, mostly, world-dependent (remember that Lewis is the proponent of the possible-worlds semantics for counterfactuals). Popperian propensities (Popper 1959) are tendencies or dispositions to produce a particular result or outcome on a specific occasion (for instance, an experiment). Gillies (2000: 125–126) characterises the propensity theory as an objective but nonfrequency theory having the following traits: (i) probability is introduced as a primitive undefined term characterised by a set of axioms; and (ii) probability is connected with observation in some manner more indirect than the definition in terms of frequencies. Gillies then distinguishes between a long-run propensity theory and a single-case propensity theory. In the former, probabilities are associated with repeatable conditions and produce, in the long run, frequencies that are approximately equal to the propensities. In the latter, propensities produce particular outcomes in specific set ups. According to Gillies, Popper held both versions. Nevertheless, Lewis' objective chance can be equated with propensities with the intent of expressing intrinsic characteristics of the world.

Chance, in Lewis' proposal is not used to *shape* subjective probabilities—it is entirely and uniquely determined once our *credence* in the truth of the corresponding proposition is fixed and evidence does not contradict this evidence. Simply put, if our credence in A is x, the chance of A is simply x.

The way in which the Principal Principle ties the objective and the subjective sides of Janus-probability differs from F-D approaches. Although Lewis accepts and indeed supports the distinction between two different concepts of probability—one subjective and one objective—he opts instead for the opposite combination. He says (Lewis 1986: 83–84):

> Carnap did well to distinguish two concepts of probability, insisting that both were legitimate and useful and that neither was at fault because it was not the other. I do not think Carnap chose quite the right two concepts, however. In place of his 'degree of confirmation' I would put credence or degree of belief; in place of his 'relative frequency in the long run' I would put chance or propensity, understood as making sense in the single case. The division of labor between the two concepts will be little changed by these replacements. Credence is well suited to play the role of Carnap's probability$_1$, and chance to play the role of probability$_2$.

Unfortunately, it is not the case that these replacements will make little difference. Indeed, it is this very same difference between the F-D accounts and the Principal Principle that vindicates the choice of the empirically-based or objective Bayesianism. As Salmon (1988: 21) points out, according to the Principal Principle, it seems possible that different agents, with different initial credence functions, will assign different *objective* chances to the truth of A. Agreed, different agents may have different estimates of the chance of A, or different degrees of belief about the chance of A, but the Principal Principle, since it allows different assignments of objective probabilities, opens the door to arbitrariness. Frequency-driven accounts do not run this risk.

Let me refer once more to the Principal Principle to underline the substantial difference of the usage of physical and subjective probabilities in twofold causality. In the single case the aim is not to claim credence about chance, e.g. credence about the chance of the causal factor 'smoking' to produce lung cancer in Harry. Rather, it is a rational degree of belief in the hypothesis that Harry's smoking caused him to develop cancer, given the available evidence about the generic causal claim. In other words, the support of the hypothesis in the single case is based on knowledge about frequencies that hold at the generic level: on this point see also later section 6.2.2. Let us now discuss some possible objections.

5.3.4 Objections

Is causal knowledge given up? The first objection comes from the staunch objectivist and the persuaded realist. Despite the differences between Bayesianisms, and despite my favouring a frequency-driven version, it is a matter of fact that Bayesian interpretations of probability deal with rational degrees of belief. But rational degrees of belief, recall, are features of an agent's mental state, i.e. they are in sharp

contraposition with physical probabilities. So, here is the threat: does the adoption of a subjectivist perspective lead to dropping the ambition to acquire knowledge about the world, notably, about causal relations? My reply, in a nutshell, is: no, as long as subjective probabilities are based on the available evidence.

Let me argue more widely. To understand why Bayesian approaches do not necessarily lead to an antirealist position concerning causal relations let us have a look again at the Carnapian *c*-function. As mentioned above, *q* is the degree to which a piece of evidence *e* supports or confirms a given hypothesis *h*. The concept of degree of confirmation can be lodged within the subjectivist framework because it is more akin to a feature of an agent's mental state rather than an objective feature of the world. Now, I would like to ask: is everything subjective in the *c*-function? What exactly does the evidence *e* state? Evidence *e* represents the experimental or observational evidence, in other words, what we know about the world, and what we shape subjective probabilities upon. Differently put, subjective probabilities are not devoid of empirical content as long as they are dependent on empirical constraints.

Are frequencies a mere pedagogical need? The second objection, instead, is raised by the staunch subjectivist. Although she would rejoice at the adoption of rational degrees of belief, she would also argue that there is a subjectivist approach that does not employ frequencies at all: this is the approach proposed by de Finetti (1937) and Savage (1954), better known as rational decision-making. Thus, she might wonder why other Bayesians—namely, objective or empirically-based Bayesians—need to introduce frequencies in their accounts. Isn't it just a pedagogical need?

Let us consider, for instance, economic contexts. The force of the rational decision-making approach seems to lie in the fact that only one concept of probability—the subjective one—is employed, and that subjective probabilities are sufficient for one's decision. Without going through the technicalities of rational decision making, the very basic idea turns on the concept of utility,[3] and the general rule for decisions prescribes the maximization of the expected utility. That is, this rule says to choose actions for which the estimate of the resulted utility has its maximum. Rational decision-making seems to be applicable even in cases in which the agent does not know beforehand the values of probability—i.e. the relative frequencies—for some events, precisely because she can make subjective assignments. That is to say, priors can be assigned arbitrarily (as long as coherence is preserved) and we can, therefore, get rid of frequencies. Thus Savage's and de Finetti's stance.

Indeed, Carnap (1951: § 50–51) is aware of the problem, and his solution seems quite accurate. In rational decision-making we can do without frequencies, he argues, if inductive logic is accepted. This is for two reasons. The first concerns the interpretation of probability$_1$. Carnap can accept that practical decisions be made upon sole knowledge of probability$_1$, because probability$_1$ can be interpreted as estimate of probability$_2$ (§ 41D). And it is true that if the agent actually knows the value

[3] In economic contexts, 'utility' measures the degree of satisfaction gained from consuming commodities, i.e. goods and services. Once this measure is specified, one can talk of increasing or decreasing utility, and explain economic behaviour in terms of the attempt to maximise one's utility.

of probability$_2$, then the corresponding value of probability$_1$, with respect to this evidence, simply equals the value of probability$_2$, namely the known relative frequency (§ 41C). Nevertheless, the problem still remains, because values of probability$_2$ are unknown in the great majority of cases concerning, for instance, ordinary economic decisions.

And here comes the second reason. Values of probability$_1$ are not said to be unknown in the same sense in which probability$_2$ are. That a certain probability$_2$ value is unknown means that we do not have sufficient factual information for its calculation. Whereas a probability$_1$ value cannot be unknown in the same sense; according to Carnap, a probability$_1$ value is unknown in the sense that a certain logico-mathematical procedure has not been performed yet. The most common situation is that the only available information concerns the frequency of a property M pertaining to an observed sample, hence the relative frequency of M in the whole population is unknown. Nonetheless, what is still possible is to calculate the probability$_1$ of a hypothesis which ascribes M to an unobserved individual. This probability$_1$ value is the estimate of the unknown probability$_2$ at stake. And this probability$_1$ value will be enough as a basis for the agent's decision. So to go back to the subjectivist's question, do we really need frequencies? It seems that, even though subjective probabilities may be a sufficient basis for decision-making, in the ultimate analysis those subjective values are formed upon frequencies, pace de Finetti and Savage. So far, so good.

Exchangeability. Nonetheless, the staunch subjectivist can still play her last card: *exchangeability*. De Finetti's exchangeability argument (de Finettti 1937) claims that all probabilities are subjective, and that even apparently objective probabilities can be explicated in terms of degrees of belief. Briefly and informally, a sequence of random variables $X_1, \ldots, X_n$ is exchangeable if, for any fixed n, the new sequence $X_{i1}, X_{i2}, \ldots, X_{in}$ has the same joint distribution no matter how $i1, \ldots, in$ are chosen.

Let Y_n be the average of any n of the random variable X_i, namely:

$$Y_n = \frac{X_{i1} + X_{i2} + \cdots + X_{in}}{n} \tag{5.1}$$

Since sequences are exchangeable, it does not matter which sequence $i1, i2, \ldots, in$ is chosen. De Finetti then shows that the different distributions thus generated will tend to a limit as n tends to infinity. That is to say, different agents may start with different prior probabilities, but, as evidence is accumulated, their posterior probabilities will tend to converge, thus giving the illusory impression that objective probability exists; therefore, de Finetti interprets his mathematical result as showing that we can get rid of objective (or physical) probability. In his view, objective probability is metaphysical in character, and the exchangeability argument could be accepted if one wanted to raise doubts against metaphysical propensities.

Exchangeability might be challenged, however. For instance, Gillies (2000: 77 ff.) takes de Finetti's argument as a reduction from the objective notions of (objective) probability and independence,[4] to the subjective notions of (subjective)

[4] If two events E and F are independent then the joint probability $P(E \,\&\, F)$ is equal to their product, i.e. $P(E) \times P(F)$.

probability and exchangeability. He then argues against this reduction by saying that the concept of exchangeability is actually parasitic on that of objective independence and, consequently, redundant. Let me explain this point more thoroughly.

As we have seen earlier, according to the subjective interpretation, different agents, although perfectly reasonable and having the same evidence, may have different degrees of belief in the same event or hypothesis and the mathematical theory of probability provides a tool to measure those degrees of belief. However, next to subjective probabilities there seem to be objective probabilities too; for instance, the probability that an unbiased dice will show an even side or the probability that a particular isotope of uranium will decay in a year seem to be objective rather than subjective. De Finetti's exchangeability argument is meant to show that even these probabilities are apparently objective, apparently because even if agents adopt a strictly subjectivist view, as long as they update prior probability values by Bayesian conditionalisation, they will come to an agreement on their posterior. Gillies' critique of exchangeability (Gillies 2000: 69 ff.) involves two lines of argument. The first one questions the use of Bayesian conditionalisation and can be summarised as follows.

Prior probability functions will in all cases be based on general assumptions about the nature of the situation under study. If these assumptions are correct, then updating prior probabilities by Bayesian conditionalisation will yield reasonable results. However, if these assumptions are wrong, the prior probability values as well as the posterior will be inappropriate. In this case we will not just update the priors but we will need to choose drastically new prior probabilities values for the event or hypothesis. Therefore, allowing changes in probability values *only* by Bayesian conditionalisation as de Finetti does is just too restrictive.

The second line of argument questions the general assumptions on which prior probabilities values are based. Gillies (2000: 75–77) claims that the subjective concept of exchangeability is equivalent to the objective concept of independence. In a strictly subjectivist framework such as de Finetti's, we can get rid of independence. However, says Gillies, from an objectivist viewpoint we can apply exchangeability only in a case of independence. The trouble is that there are many situations in which the outcome of an event is strongly dependent on the outcome of a previous event and here exchangeability will deliver erroneous results. Gillies concludes that we are not able to reduce independence to exchangeability, but we can reduce exchangeability to independence. That is to say, to use exchangeability safely, we have to know first that we are in a situation of independence.

It seems to me that exchangeability is not, after all, a decisive argument against the objective or empirically-based Bayesian interpretation, and particularly against an epistemic use of frequencies. Empirically-based and objective Bayesianism do not 'reify' objective probabilities into metaphysical propensities. Objective probabilities have a preferred frequentist interpretation, and there is nothing 'untestable' in this. Relative frequencies are instead known by experience, and this is what guarantees that even subjective probabilities shaped upon them are not devoid of empirical content. Quite apart from Gillies' critique, the idea that in exchangeable sequences epistemic probabilities are determined by the corresponding frequencies is reiterated also in Courgeau (2007b), where he discusses statistical inference from a

population to a new individual in a standard logit model. For a large n, the epistemic probability that the event A will occur in the $(n+1)$ case, simply is the frequency of A in the previous n cases. This obviously presupposes that we are dealing with exchangeable sequences, but the point at stake is that epistemic probabilities are frequency-driven and that this result holds for standard logit models and can be extended to multilevels too. The game is still open.

The subjective Bayesian will rebut again that frequencies are just a pedagogical tool, and that subjective Bayesianism, in ultimate analysis, is not ruled out in principle. Different scientists can allocate priors in different ways and all be equally rational: coherence, according to Dutch book arguments, is also a sufficient condition. I might concede that subjective Bayesianism is not ruled out in principle, although in this way subjective Bayesianism makes sense of 'learning from experience'—which, recall, is one of the two motivations for the Bayesian framework—only after priors have been allocated. Let us ask: does experience teach us anything at all before the allocation of priors? It does. And this is why we'd better let our rational degrees of belief be frequency-driven. Conversely, if experience does not teach us anything about actual frequencies, for instance if they are not available, according to the logical constraints of the objective Bayesian account, we have to set the degree of belief to 0.5. But in this case, experience will still have taught us to use a middling value.

5.4 Objective Bayesianism in Causal Modelling

5.4.1 Empirically-Based vs. Objective Bayesianism

In the previous section, I offered an argument to support the empirically-based and the objective Bayesian interpretations. The argument relied and appealed to *theoretical* issues: the need (i) to account for different types of probabilistic causal claims and (ii) to make sense of learning from experience. This section goes a step forward and evaluates the methodological advantage of adopting a Bayesian, and particularly the *objective Bayesian* interpretation. Specifically, I shall consider two aspects: (i) design and interpretation of tests, and (ii) guide for action. These two tasks are related to the *practice* of social research, and the goal is to show that the interpretation of probability is not neutral with respect to it; that is to say, interpreting probability is not an academic debate for armchair philosophers, but has important consequences for the methodology and practice of causal modelling. The previous section already isolated two viable Bayesian interpretations, so we will have now to choose between the empirically-based and the objective Bayesian accounts.

The choice lies in the constraints to set single-case probabilities. Under the empirically-based account, single-case probabilities are shaped upon knowledge of frequencies. However, if such knowledge is lacking, the agent is free to choose any probability value. Different agents can choose different probability values and all be rational, as long as coherence is preserved. Somehow, in lack of empirical

constraints, empirically-based Bayesianism collapses into subjective Bayesianism. This is of course a dangerous situation, for subjective Bayesianism is arguably too arbitrary a position to hold in science. Objective Bayesianism, on the contrary, does not fall short of arbitrariness because in absence of any empirical information, degrees of belief should be as equivocal as possible. The objective Bayesian is always forced to choose the probability value that is closer to the middling value. Thus different agents setting different probability values to the same events will not be deemed equally rational and there will be very little room for arbitrariness. Williamson (2007) makes the case for objective Bayesianism over the empirically-based account on these grounds, and Russo and Williamson (2007b) apply this result to the domain of cancer epidemiology. In the following, the same line of argument will be adopted for causal modelling in the social sciences.

5.4.2 Design and Interpretation of Tests

In section 3.3, I already pointed to the interpretation of tests as a major difficulty in causal modelling. There, I especially emphasised the choice of the test statistics, the threshold of the significance level, and the somehow misleading meaning of 'significant'. I also mentioned that, from a frequentist point of view, we cannot test the probability of a hypothesis being true, but only the probability of obtaining the observed sample *if* the hypothesis is true. This last issue is indeed a crucial one if we want to assess the methodological gain in adopting a Bayesian interpretation rather than a frequentist one, as is customarily done in causal modelling.

In the frequentist interpretation, as well as in Kolmogorov's axiomatisation, probability is defined over sets. However, probability can also be defined over propositions. Although an equivalence between those two semantics can be established, we cannot always consider a proposition as the set of the disjoint subpropositions. That is exactly the case in hypothesis testing. Thus, if we want to test a hypothesis, that is a proposition to which we attach a *single-case* probability value, the Bayesian framework will do a better job than frequentism.

To begin with, when we say that an unknown parameter θ lies in the interval (θ_1, θ_2) with a confidence level of 95%, this gives the erroneous impression that the parameter θ has a certain probability of lying in the specified interval. Instead, this means that if we draw many samples of the same size and build the same interval around θ, then we can expect that 95% of the confidence intervals will contain the unknown parameter. Needless to say, this is *not* the same thing as asking what the probability that a given parameter will lie in a given interval is (on this point see Courgeau 2004b). For this reason Freedman et al. (1998: 347), discussing confidence intervals and the frequency interpretation, say that "chances are in the sampling procedure, not in the parameter". The frequentist is thus unable to answer those kind of questions exactly because frequencies represent the percentage of occurrences of an event. No matter how many samples we draw and how many confidence intervals we build, the question of whether the parameter will lie in the

specified interval remains unanswered—single-case probabilities do not make sense in the frequentist account. But the Bayesian does have an answer to this question because she can attach probability values to a single instance, that is to the hypothesis that the parameter will lie in the interval. Jaynes made a similar point as early as 1976 (see Jaynes 1989: ch. 9). In this paper he analysed six common statistical problems and showed that Bayesian methods (i) are easier to apply, and (ii) yield the same or even better results. Jaynes is not an isolated case, though, as in the Bayesian statistical literature many similar attempts can be found. See for instance Florens and Mouchart (1989, 1993), Drèze and Mouchart (1990), and, from a more specific *objective* Bayesian viewpoint, Berger et al. (1997), Berger and Delampady (1987), Bernardo (1980).

But there is another reason why we should favour the objective Bayesian account.[5] Hypothesis testing typically tests the null hypothesis against the alternative hypothesis. Somehow there isn't an equal treatment of the competing hypotheses, because acceptance or rejection will *directly* concern the null hypothesis and *only indirectly* the alternative hypothesis. On the contrary, under the objective Bayesian approach both hypotheses are treated as being equally probable, that is the objective Bayesian does not favour the null hypothesis over the alternative, unless there is evidence for this choice.

The frequentist will probably rebut that there is indeed a serious reason for testing the null hypothesis against the alternative hypothesis, and not the other way round. The reason lies in the graveness of committing a type I or a type II error. A type I error is made if the null hypothesis is rejected when in fact it is true, and a type II error is made when the null hypothesis is accepted when instead the alternative is true. A type I error is said to have probability α, and the value of α is also called the level of test, or better, the p-value. Type II error is instead said to have probability β. It is commonly agreed that committing a type II error is by far weightier than committing a type I error. In other words, we should be more cautious in accepting the hypothesis that the detected variation is true (alternative hypothesis) rather than the hypothesis that the detected variation is chancy (null hypothesis).

The first thing worth noting is that whilst α has a standard value given by the significance level, the computation of β is less than obvious. β depends but is not determined by α: the two measures are related but it is not possible to compute one from the other (on this point, see for instance Bayarri and Berger (2000) and Thompson (2006)). Moreover, the result of the test in the frequentist framework depends on the probability of error because the decision will depend on the rejection region. Restricting the rejection region of α will reduce the probability of a type I error, but what about *the probability of the alternative hypothesis*?

Let us see what would happen in a Bayesian and particularly in an objective Bayesian framework. First of all, as just said, the objective Bayesian will treat both hypotheses as equally important and assign a higher weight to the null hypothesis only if the evidence suggests it. In the Bayesian framework, the problem is

[5] I wish to thank Jon Williamson for drawing my attention to this point and Juan Carlos Martinez-Ovando for helping to shape the argument.

not to find a rejection region, but to work out the posterior probability of the two hypotheses and to make a decision on this basis. All information needed to make the decision is contained in the posterior. This is typically done using Bayes factors; Bayes factors give the posterior odds of a hypothesis when both hypotheses have equal priors. For one account of Bayesian hypothesis testing, see for instance Berger and Guglielmi (2001). The risk of committing a less or weightier error has now disappeared. More to the point, in the objective Bayesian framework we have a meaningful way of evaluating the *probability of a hypothesis*, and the decision of accepting or rejecting a hypothesis is based on empirical evidence.

5.4.3 Guide for Action

Famously, the Bishop Butler said that *probability is the very guide of life*. That is to say, probability has to guide our actions, it has to inform our decisions. This raises at least two questions: What kind of decisions? And, what kind of probabilities can inform our decisions?

Let us consider types of decision first. As far as causal modelling is concerned, there are two. The first type relates to the cognitive goal of causal modelling and the second to its action-oriented goal. Decisions pertaining to the cognitive goal have already been discussed above. Objective Bayesian probabilities allow us to decide whether to accept or reject a hypothesis by comparing the posterior probabilities of competing alternatives. I shall now focus on the action-oriented goal. Decisions and actions can have two objectives: policy-making and decisions about particular individuals (for brevity, let us call them 'single-case decisions'). Arguably, the social sciences are concerned more with the former than with the latter. Conversely, other sciences, e.g. the health sciences, are primarily concerned with single-case decisions.

Let us first clarify a couple of issues. I will then come back to the type of probability that is best suited to guide actions. Firstly, although demography might not be specifically interested in Harry's probability of migrating, governmental policies have to have a bearing on the individual, because they are meant to induce changes in the population's migration behaviour, that is on the behaviour of individuals. The same holds for policies concerning public health. Secondly, although a physician might be interested in the probability that Harry, who has lung cancer, will survive more than two years, this single-case probability is (partly) computed using population frequencies. Therefore, although policy-making and single-case decisions are quite distinct things, they are indeed related.

We now come to the kind of probabilities that have to inform decisions. The reason why single-case decisions should be taken on the basis of objective Bayesian probabilities is twofold: (i) these probabilities are applicable in the single case—on the contrary, the frequency interpretation does not make sense in the single case—and (ii) objective Bayesianism leaves no room for arbitrariness. It can also be argued that objective Bayesianism leads, on average, to more cautious actions in the case

of risky decisions—see for instance Williamson (2007) and Russo and Williamson (2007b) on this point.

But why policy making is better off with an objective Bayesian account? Let us read the Introduction of the *Policy Hub* of the UK Civil Service (http://www.policyhub.gov.uk/better_policy_making/, accessed 1 June 2007):

> Policy making is: "the process by which governments translate their political vision into programmes and actions to deliver 'outcomes'—desired changes in the real world". (Modernising Government White Paper, 1999) This concern with achieving real changes in people's lives is reflected in the Government's overall strategy for improving public services published in March 2002 (Reforming our public services: principles into practice) Promoting good practice in policy making is fundamental to the delivery of quality outcomes for citizens and to the realisation of public sector reform. Policy makers should have available to them the widest and latest information on research and best practice and all decisions should be demonstrably rooted in this knowledge.

This quote reveals two fundamental elements suggesting the desirability of opting for the objective Bayesian approach. To begin with, policies intend to induce changes in the world. This is important to justify the *causal* interpretation of models—if statistical models merely *described* phenomena, it would be useless to set up policies based on results of research. Also, if the results of research merely stated frequency of occurrence, we would not have any grip on the causal relation—what we need is to form an opinion, i.e. an epistemic attitude, about the causal relation at stake that can trigger actions and decisions. Bayesianism offers such a tool—degrees of belief—and objective Bayesianism offers normative precepts that avoid arbitrariness to determine those degrees of belief. Although in Bayesian interpretations probabilities are quantitative expressions of an agent's mental state, I argued in section 5.3 that degrees of belief are not devoid of empirical content insofar as they are shaped upon available evidence. Objective Bayesianism goes even a step forward in advising on how to set degrees of belief when evidence is lacking. A second element is that policies ought to be decided on the basis of the widest available knowledge. Objective Bayesianism is again the best candidate because it allows us to take into account *any* type of evidence. Knowledge of frequencies, for instance, belongs to this evidence. Results of associational studies will then acquire importance because they can be evidence to be incorporated in causal modelling.

The point is not to dismiss classical statistics in favour of Bayesian statistics. Bayesianism is an epistemological stance about reasoning and inference in science. But the epistemological stance we adopt should also (positively) influence our scientific practice. The objective Bayesian practising scientist will be in a better position to build, test and interpret causal models, let alone advice policy making and single-case decisions.

Chapter 6
Methodological Consequences: Mechanisms and Levels of Causation

Abstract This chapter opens by offering an account of the notion of mechanism in causal modelling. It is argued that causal modelling ought to be the modelling of mechanisms and that, by modelling mechanisms, causal models are able to provide explanations of social phenomena. The chapter then discusses the problem of the levels of causation. It first reformulates Sober's Connecting Principle by which we can calculate the support of a single-case causal hypothesis on the basis of generic causal knowledge. Afterwards, the chapter presents a taxonomy of the types of variable and in fallacy of causal models, and argues that the problem of the levels of causation is better understood and dealt with as a problem of levels of analysis.

Keywords Mechanisms; explanation; hypothetico-deductive explanation; twofold causality; levels of causation; levels of analysis; connecting principle; ecological and atomistic fallacies.

Introduction

Chapter 4 developed at length the rationale of variation. I explained how this rationale works in causal modelling and later, in Chapter 7, I will further support my proposal by showing that the rationale is compatible with a number of accounts proposed in the philosophical literature. This chapter discusses other methodological consequences of the epistemology of Chapter 4: it attempts to give an account of the notion of mechanism and of the problem of the levels of causation.

Section 6.1 discusses the notion of *mechanism* and offers an understanding of what mechanisms are in the context of causal modelling. Attention is also drawn to the modelling of mixed mechanisms, to the difference between modelling mechanisms vs. modelling decision-making processes, and to the explanatory import of mechanisms in causal models. The distinction between modelling mechanisms and modelling decision-making processes raises the problem of the level of causation, which is dealt with in the second section.

F. Russo, *Causality and Casual Modelling in the Social Sciences: Measuring Variations* 155
Methodological Prospects in the Social Sciences

Section 6.2 first presents and questions the plausibility of metaphysical accounts of the levels of causation; it then offers an account of the levels in line with causal modelling. The goal here will be threefold. First, to unveil a fallacious way of reasoning about the levels of causation; second, to provide a better epistemological understanding of the levels of causation; and third, to formulate a principle capable of connecting the levels of causation. Further, I argue that the problem of the levels of causation is better understood and dealt with if reformulated as a problem of levels of analysis, to which Section 6.3 is devoted.

In Section 6.3, I offer a taxonomy of the types of variable used in causal modelling and an overview of the possible fallacies that occur in it. These are necessary ingredients in order to give a comprehensive account of the levels of analysis. In particular, I attempt a reformulation of the metaphysical problem of the levels of causation as a methodological problem related to levels of analysis. I then discuss the extent to which multilevel analysis and the rationale of variation succeed in accounting for the reformulated problem.

6.1 Mechanisms

6.1.1 Modelling Mechanisms

As I mentioned in Chapter 1, the term *mechanism* occurs very often in the causal parlance of social scientists. As a matter of fact, it occurs very often in the causal parlance of philosophers too. I also mentioned that 'causal link', 'causal path', 'causal chains', etc. are typically used interchangeably. I then raised the question whether this practice be correct and I suggested that those terms are not completely equivalent. In particular, I suggested that 'causal link', 'causal path', or 'causal chain' are not completely equivalent to 'causal mechanism' in at least two cases. Firstly, when more than two variables are involved, and secondly, when variables involved are of a different sort (social, demographic, economic, biological, etc.). I also suggested that 'mechanism', in spite of assonance, does not rhyme with mechanist but with *structure*, the web of relationships, and that even mechanistic accounts of causality, such as Salmon's account discussed at the beginning of Chapter 1 (but see also later section 7.1), do not necessarily commit one to deterministic causal relations; nonetheless, from the first epistemological and methodological morals drawn at the end of Chapter 1, we learnt that, in quantitative social science, causal mechanisms are *statistically modelled*. It seems to me that it is on this ground that an account of mechanisms ought to be given. In the following, I discuss some accounts that can be found either in the philosophical literature or in the scientific literature in order to develop a concept of mechanism that suits causal modelling. I will then draw a distinction between (i) evoking causal mechanisms, and (ii) modelling causal mechanisms.

Daniel Little, in his entry 'causal mechanisms' for the *Encyclopedia of Social Science Research Methods*, opposes Humean causality, that sees causation as mere regularity, to a realist view, that sees causal mechanisms and causal powers as fundamental. According to the realist, says Little (2004), "a mechanism is a sequence

of events or conditions governed by lawlike regularities leading from the explanans to the explanandum". Partisans of this view are obviously Salmon, for he believes (i) that causal processes, interactions and laws give the causal mechanisms by which the world works, and (ii) that to understand why something happens we have to disclose the mechanism that brought it about (Salmon 1984: 132); and Dupré and Cartwright (1988). who, along the same lines, argue that discovering causal relations require substantial knowledge of the capacities of things or events—i.e. their power to bring about effects. It is worth noting that the exact role capacities play in causal mechanisms is not sufficiently spelled out, though. More to the point, the index of Cartwright's 1989 influential book even misses the entries 'mechanism' and 'causal mechanism'. Apart from reference to causal powers, the only hint we have is that a mechanism is a 'sequence of events'.

Unfortunately, we do not go much further by reading Patrick Suppes. Suppes (1970: 82) states that mechanisms are "the means by which causes operate", which would be helpful if he provided us with an explication of what these 'means' are. But he doesn't. He instead lingers on the role that the concept of mechanism played in the history of thought. The concept of mechanism is related to that of ultimate/remote cause, that is the sort of causes to which all other causes should be reduced.

Woodward (2003: 48–49, 52–53), discussing the issue of modularity of equations, says that, in a system, each equation represents a distinct and independent causal mechanism. Informally, given a set of equations, the system is modular if it is possible to set up interventions (i.e. to modify the value of the dependent variable or of its probability distribution) in any one of the equations without 'disturbing' any of the other equations. If the system has the modularity property, then each equation represents an independent mechanism upon which we can intervene. Again, mechanism is a synonym of "means by which *A* affects *B*" (Woodward 2003: 3), but it is not further explicated.

Aish-Van Vaerenbergh (2002: 52), in a paper on explanatory models in suicide research, makes the following claim about mechanisms, which is worth reading in full:

> Clarifying why and how certain factors are related to suicidal behaviour is a matter of explaining the *causal mechanism* underlying such behaviour. It is an attempt to specify the *causal chain* through which the antecedent (the cause) has an influence on an outcome (effect), or, in Baron and Kenny's words (1986: 1173),[1] an attempt to specify "the *generative mechanism* through which a focal variable is able to influence a dependent variable of interest". (My emphases.)

In the context of her paper, this claim is lucid. Aish has a practice-oriented perspective—her interest being social policy. Her research in suicide behaviour is meant to orient prevention and intervention programs that are designed to change critical mediating variables (the causes) thought to be causally related to the outcome variable and thus bringing about a change in the outcome (suicide behaviour). This explains her interest in mechanisms *as means* through which causes operate, or through which *A* affects *B*, etc. According to Aish the causal mechanism is "the

[1] Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator distinction in social psychological research: conceptual strategic and statistical considerations. *Journal of Personality and Social Psychological*, 51(6), 1173–1182. Quoted in Aish-Van Vaerenbergh (2002: 52).

causal chain through which the antecedent (the cause) has an influence on an outcome (effect)" and then "the generative mechanism through which a focal variable is able to influence a dependent variable of interest". However, isn't Aish mixing up too many notions—causal mechanism, causal chain, generative mechanism—together? How are these notions articulated? How do they hinge upon one another?

Reference to the 'generative mechanism' comes into range at the right moment because it allows us to highlight a number of important points. Firstly, Aish claims that the *mechanism* is 'the means by which causes operate'. Secondly, and most importantly, she adds that the mechanism is *generative*: it *generates* the effect; it causally *operates* on the effects as well as on the causes. Thirdly, the mechanism specifies the *causal chain* from the cause to the effect. Now, depending on how the notion of 'causal chain' is understood, Aish's account can shed light on the notion of mechanism. Notably, if she simply means a temporal sequence of cause-and-effect events, then, perhaps, this does not take us much further because we lose the complexity of causal structures. Yet, this is already an improvement with respect to accounts that consider causal relations, quite simplistically, as relations between an event-cause and an event-effect. In Aish's chains we have intermediate variables—the 'means' by which causes produce their effect.

But the conceptualisation of the notions of mechanism and of generative mechanism can be pushed further. Franck (2002), notably, makes a significant step in this direction. In a nutshell, the generative mechanism *is not* simply a causal chain, but a *system* in which causal variables *and* functions act. Franck developed the notion of generative mechanism as part of a vast collective research project—the first volume of the *Methodos Series*—to which also Aish's contribution belongs. Thus Aish, Franck, and the other authors of that volume synergetically come to define modelling strategies applicable to different domains and aiming to increase the explanatory power of models.

Let us now analyse Franck's position. To illustrate, Franck uses McCulloch and Pitts' model of a neuron. A neuron is a generative mechanism, *generating* an effect when it is stimulated by some causes. The model of the neuron—i.e. the picture of the neuron, so to speak—represents its functional architecture, but it *does not* represent the material structure, i.e. the soma, the dendrites, etc. Consider the model representing just one neuron as in Fig. 6.1 (Franck 2002: 143). In this picture five functions are represented: (i) the reception of stimuli, that is x_1, x_2,…, x_N; (ii) the weighting of stimuli, that is the synaptic coefficients w_1, w_2,…, w_N; (iii) the sum p of the weighted stimuli; (iv) the threshold of stimulation σ, above which transmission occurs; and (v) the stimulus exits. This model, as Franck says, represents the functions themselves. Such a model is wholly conceptual, that is formal or theoretical, and yet real since it governs the mechanism. The two structures of the neuron, the functional one and the material one, could both be represented within a single model, but we can also represent the functional structure of the generative mechanism by means of a model from which all material content has been removed, as shown in McCulloch and Pitt's model. Thus, modelling the *material* architecture of the neuron and modelling its *functional* architecture, i.e. the combination of *functions* which govern the generative mechanism are to be kept distinct.

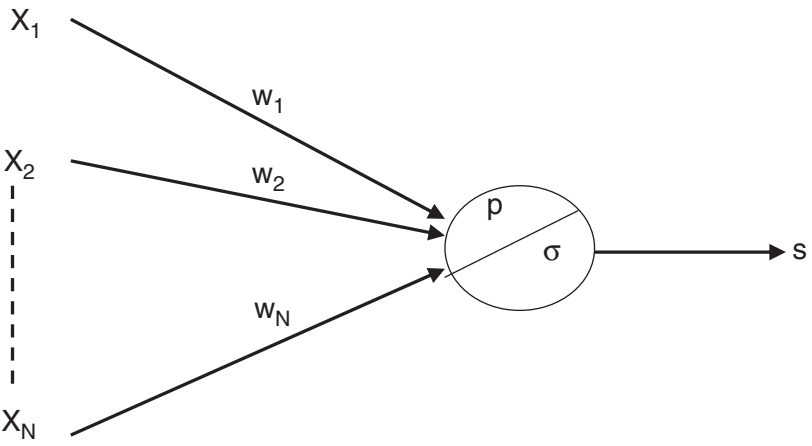

Fig. 6.1 McCulloch and Pitts' model of a neuron

Let us now apply to a *social* mechanism the distinction just made for the neuron between modelling its material and its functional structure. Where does causality come in? Causality is in the combination of social factors (i.e. the causes) operating in the social mechanism. Modelling the combination of social factors of this mechanism concerns the material architecture of it, not its functional architecture. Franck claims that we need to model the functional structure of the social mechanism as well as its causal structure. Functional modelling of the social generative mechanism is thus complementary to the causal modelling of this mechanism. Functional modelling is complementary because modelling the functions of a social generative mechanism allows us to select and to group together relevant variables in the causal model of the generative mechanism, according to the functions that the causal variables carry out in this mechanism. That is to say, variables in a causal model ought to be selected and grouped together, according to Franck, depending on the role (i.e. the function) they play in the generative mechanism.

To illustrate, consider the metaphor of the black box. The box represents the generative mechanism, which is both a causal and a functional system. The effects generated by the social generative mechanism are the properties of this mechanism. Effects or properties of the mechanism are the outputs (Y) of the box when certain inputs (X) occur (see Fig. 6.2). The input is given by (some of) the causes which are, like outputs, external to the box. Functions of the generative mechanism are inferred from the study of its effects or properties, which are the outputs of the box. Since the

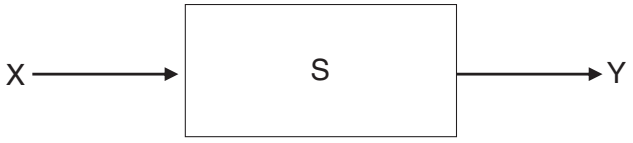

Fig. 6.2 A black box with inputs X and outputs Y of a system S

properties of the mechanism are external to it, the functions of the mechanism may be discovered without opening the box, i.e. without knowing anything about the way causal factors operate within the generative mechanism. On the contrary, in order to know the way causal factors operate, we ought to open the box, which can be done by causal modelling. However, according to Franck, once the functional structure of the mechanism has been inferred from the study of its properties, causal modelling may be founded upon it, since the functional structure allows us to select, among many social factors which could contribute to the effects, those factors that carry out the functions without which the mechanism would not *generate* these effects.

The generative mechanism, recall, is 'the means by which causes operate'. Now, because the generative mechanism is disentangled by using causal models (for instance those described in Chapter 3), in Franck's account 'the means by which the causes operate' is indeed spelled out. Sections 6.1.2 and 6.1.3 below develop more on this idea—that causal modelling is, and ought to be, the modelling of mechanisms. But functional modelling is meant to be complementary to causal modelling in the sense that modelling the functions will guide the causal modelling practice. Developing further Franck's ideas about functional modelling is indeed an interesting and promising path of research, as shown in Franck (2002).

Recently, Mario Bunge (2004) has given a rather comprehensive story about mechanisms. Functions, systems, black boxes, and mechanisms themselves enter Bunge's account in a different way than Franck's. The main message Bunge puts forward is that to explain a fact is to exhibit the mechanism that makes the system in question tick. There is a first patent difference between Bunge's approach and mine. Whilst Bunge is asking what makes an explanation good—and the answer is: the exhibition of mechanisms—I am rather interested in understanding what it is to *model* a mechanism. Yet Bunge is very argumentative and our discussion will definitively benefit from reading him. The core of Bunge's account turns around the definitions of *mechanism* and of *system* (Bunge 2004: 186–188):

> In all of the above scientific examples, a **mechanism** was conceived of as a *process (or sequence of states, or pathways) in a concrete system, natural or social*. […] A **system** is a complex object whose parts and components are held together by bonds of some kind. (Emphases in the original, my bold.)

Bonds, according to Bunge, can be logical or material. Furthermore, depending on the bonds and on the system's constituents, a concrete or material system may belong to different levels: physical, chemical, biological, social and technological. It is worth noting that here 'system' is not a synonym for 'structure' nor for 'black boxes with inputs and outputs' (Bunge 2004: 187), but is intended as a set of elements standing in reciprocal interrelations and mathematically defined by certain families of differential equations—this is the notion of system developed in system analysis or systemics (see von Bertalanffy 1969 and Bunge 1979b). Mechanisms are, in the ultimate analysis, reduced to processes; mechanisms (i.e. processes) can interconnect as is the case in most biosystems or social systems.

A couple of remarks. Firstly, this reduction of mechanism to process looks like passing the buck and, furthermore, Bunge fails to specify whether processes have to be understood in terms of *physical* processes *à la* Salmon or in a different way.

If this is the case, the plausibility of physical processes in social contexts has already been challenged in Chapter 1. Secondly, in Bunge's account, mechanisms have more the appearance of linear (possibly, also temporally ordered) chains and the whole complexity of the 'gear', i.e. of the interconnections of the components of the mechanism, is deferred to the system. Systemism, says Bunge, is an appealing approach to anyone whose endeavour is to explain the formation, maintenance, repair, or dismantling of a concrete complex thing of any kind (Bunge 2004: 190).

Another important feature of Bunge's mechanisms is their being governed by causal laws. Not all mechanisms are causal or deterministic in the strict sense—there are also stochastic or random processes—but *causal* mechanisms are *lawful* mechanisms. The scientific plausibility of mechanisms comes from their being lawful, or otherwise they are miracles. However, says Bunge, in the context of the social sciences, laws and mechanisms are insufficient, yet necessary, to explain "because almost everything social is made rather than found" (Bunge 2004: 197). I would like to challenge this last claim.

If lawful mechanisms are necessary but insufficient to provide explanations in the social realm, it is because the received concept of law is not applicable to social contexts. The received concept of law is saturated with regularity, necessity, and universality. But the objects of the social sciences are too variable and mutable to be chained into universal regularities. The social sciences will be better off by abandoning the regularity framework and switching to a framework of *variation*, where social laws are replaced with empirical generalisations that describe causal variations between variables of interest and that allow outlying observations and non-instantiated cases.

I will now sketch the features of causal mechanisms in a much simpler manner. To this purpose, it will be useful to dwell a bit upon the concept of causal structure, which is intuitively clearer and easier to grasp. Let us read Cartwright (1997: 343):

> A causal structure is a fixed (enough) arrangements of components with stable (enough) capacities that can give rise to the kind of regular behaviour that we describe in our scientific laws.

Causal structures are graphically representable in causal diagrams. The key idea we should keep from Cartwright's definition is that of an 'arrangement of components'. The causal mechanism has to specify what this arrangement looks like, either pictorially by means of diagrams, or mathematically by means of structural equations. However, if these 'components' are identified with capacities, the mechanism stays a metaphysical entity, the intelligibility of which is guaranteed by an equally metaphysical black box that gives raise to the observed regularity. In other words, we are passing the buck to the black box without solving the problem of epistemic access to causal relations *through* mechanisms.

My point is that causal modelling is, and indeed ought to be, the *modelling of mechanisms*, and that a statistical characterisation of mechanisms, along with the variation rationale, is what mediates our epistemic access to causal relations. The net gain of this perspective is a nonmetaphysical characterisation of causal mechanisms. In fact, mechanisms have observable components (corresponding to observable variables) and the only non-observed parts of causal mechanisms are nodes

representing either conceptual or latent variables. However, far from giving causal mechanisms a mysterious or epistemically inaccessible appearance, conceptual and latent variables ought to be introduced to facilitate the interpretation of complex causal structures.

But the most important thing is that the modelling of mechanisms ought to rely on the rationale of variation rather than on the rationale of regularity. Differently put, causal mechanisms are made of *variational* relations rather than regular relations. In a causal mechanism, the components are arranged depending on what *variations* hold. Agreed, those variational relations happen to be regular (or at least regular enough), but this depends on the fact that causal modelling analyses large (enough) data sets. Furthermore, regularity does not seem to succeed in constructing causal mechanisms, for the Humean view and the realist view eventually collapse in the same tenet. In fact, according to the realist, the sequence of events in causal mechanisms is, in the ultimate analysis, governed by a lawlike *regularity*.

Let me emphasise that by saying that the causal models model mechanisms, I am not reducing the mechanism to the identification of a regression coefficient. This is a view of causal modelling reported by Hedström and Swedberg (1999b: 9–10):

> In sociology the most systematized form of black-box explanation can be found in the so-called causal modeling approach (Duncan 1975) [...]. In the causal-modeling tradition, the explanatory "mechanism" simply is a regression coefficient linking *I* and *O*, and this regression coefficient (if the model includes all relevant variables) is supposed to describe the causal influence of *I* on *O*.

Here, *I* and *O* stand for two different types of variables, that is *I*nput variables and *O*utput variables, or explanans and explanandum. *I* and *O* are linked, in Hedström and Swedberg's story, by a mechanism *M*. However, this is a very narrow conception of causal modelling. Instead, as we have seen in Chapter 3 and Chapter 4, causal models consist of a complex apparatus of statistical, extra-statistical and causal assumptions, and of background knowledge, and they employ a hypothetico-deductive methodology. It is therefore misleading and incorrect to reduce the modelling of mechanisms to the estimation of a regression coefficient.

So far I have argued that in causal modelling a causal mechanism has to be understood as an arrangement of variables (in structural equations) and nodes (in causal diagrams) that satisfies certain *variational* relations. Is this enough to claim that we are *modelling* causal mechanisms? I think it is not. To understand why, let us go back to Caldwell's study on child mortality in developing countries. By means of the following example, I intend to warn the reader about a possible confusion: *evoking* causal mechanisms and *modelling* causal mechanisms are quite distinct tasks.

A Case Study: Mother's Education and Child Survival (Caldwell's Model). Consider again Caldwell's model presented in section 6.1.3. We find here two variables, i.e. mother's education and socio-economic status, causing child survival. Strictly speaking, mother's education does not cause (or prevent) child survival, at least not in the same sense as we say that hitting the billiard ball causes it to move. Mother's education does not cause anything at all, unless we make explicit

the *causal mechanism* behind it. Caldwell (1979: 409) then explains that mother's education

> […] serves two roles: it increases skills and knowledge as well as the ability to deal with new ideas, and provides a vehicle for the import of a different culture.

The mechanism is only outlined, *evoked*; but it is not *modelled*. Caldwell's claim is a justification of why maternal education ought to be included in the model, but it does not shed light on the 'means'—i.e. the other variables—through which maternal education influences child mortality. If Caldwell's goal is to give a significant contribution to the little attempt to *explain the causal mechanism* from mother's education to child survival, then he is clearly missing the point. We are getting there though, but very slowly.

6.1.2 Mixed Mechanisms

As mentioned above, the standard characterisation of mechanisms in terms of physical process does not suit the case of the social sciences. This is for two reasons. First, if the causal model only involves socio-economic-demographic variables, we cannot identify causal mechanisms in terms of physical processes and interactions (on these grounds I also cast doubts on Salmon's approach in section 1.1). Second, if it is a mixed model involving both social and biological variables, the causal mechanism will not be able to account for the 'social' part. Let me explain these two reasons further.

One problem is that, in social contexts, mechanisms are not always, or not necessarily, made of *physical* processes and interactions. For instance, the causal mechanism modelled in the case study on health systems and mortality (López-Ríos et al. 1992) does not involve *physical* processes and interactions. Rather, this mechanism explains the behaviour of a particular social system in terms of the relations between some of its important properties. These properties, however, do no necessarily have 'physical' reality as they might just be conceptual constructs, as for instance economic and social development. Consequently, the process leading, say, from economic development to mortality through the use of sanitary infrastructures does not correspond to a *physical* process, such as billiard balls colliding, but rather is our conceptualisation and schematisation of a much more complex reality.

Another difficulty is that if causal mechanisms are governed by causal laws, it is unclear where these laws come from, and if they come from causal modelling itself, then this leads to a vicious circle. So if we want to keep a physical notion of mechanism, the price to pay is quite high—we would have to renounce causal mechanisms in the social domain. We are not forced to accept this solution, though, if we are prepared to accept a wider concept of causal mechanism, in particular one that is based on causal modelling.

The characterisation of causal mechanisms given in the previous section allows us to incorporate in the causal model both socio-demo-political variables and

biological variables. It goes without saying that pathways in such *mixed* mechanisms have to be made explicit as there is no homogeneity at the ontological level. Health variables do not cause changes in social variables (or vice-versa) *as such*. Socio-economic status influences one's health through the possibility of accessing some sanitary infrastructures, but not directly. Arguably, the social sciences are interested in identifying causal mechanisms that involve different types of variables—this interdisciplinary stance is also a perspective undertaken in epidemiology (see for instance, Susser and Susser 1996). But let me illustrate this last point more thoroughly by considering again the case of maternal education influencing child survival.

A Case Study: Mother's Education and Child Survival (Mosley and Chen's Model). As anticipated in section 1.3, a more sophisticated framework, and a much better attempt in modelling the mechanism behind the causal relation between mother's education and child survival, is given by Mosley and Chen (1984). Mosley and Chen integrate research methods proper to the social and medical sciences and their model incorporates both social and biological variables. On the one hand, research in the social sciences traditionally focuses on the associations between socio-economic status and mortality. On the other hand, medical research primarily focuses on biological processes of disease and less frequently on mortality per se. In this new framework, Mosley and Chen distinguish proximate and ultimate determinants. Ultimate determinants, i.e. the socio-economic causes, influence child mortality through proximate determinants, in this case the biological causes. The new causal mechanism is shown in Fig. 6.3 (Mosley and Chen 1984: 5).

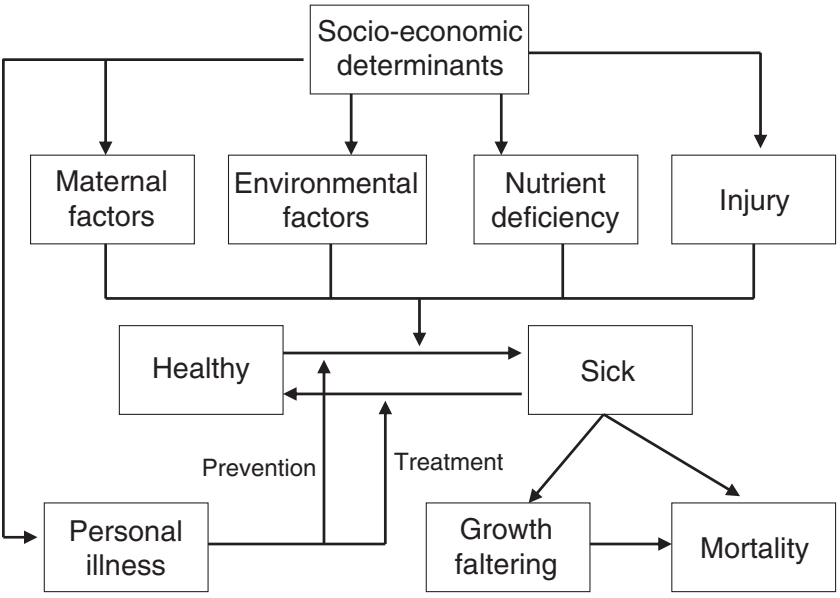

Fig. 6.3 Mother's education and child survival: Mosley and Chen's model

In this graph, a biological mechanism and a socio-economic mechanism are combined in order to explain *how* and *why* mother's education influences child survival. The answer to the how-question is given by all the possible paths in the causal mechanism. The answer to the why-question is given by some non-mathematized knowledge, viz. by background knowledge that informs both the model-building stage and the interpretation of results. Building faithful causal mechanisms is, needless to say, essential for our cognitive endeavour to understand society. Nonetheless, the action-oriented dimension cannot be underestimated. The clear and exact specification of causal mechanisms is important for policy-making purposes, since it indicates different levels at which interventions are possible or feasible. Good modelling of causal mechanisms ought to inform us about the 'means' to intervene upon.

One reason why I claimed that Mosley and Chen provide a better example of modelling the causal mechanism governing maternal education and child survival is that it incorporates both socio-demographic and biological variables. The mechanism hereby described is then a *mixed mechanism*, mixed with various types of variables. This characterisation of mixed mechanisms obviously raises a conceptual difficulty.

Strictly speaking, we should take a mechanism to be a web of relations taking place at the same ontological level. For instance, in Chapter 1, I mentioned a physiological mechanism explaining how cigarette smoking affects the lungs and possibly leads to developing lung cancer. The mechanism is here constituted by an interrelation of physical-biological-chemical processes. This kind of mechanism gives us a biomedical explanation and only involves biological variables. The mechanism in this narrow sense, however, is not the end of the story, and in particular is not enough if we need a wider view on smoking and cancer. For instance, we might need to know what parts of the population are more serious smokers in order to set up effective policies to discourage smoking behaviour and consequently reduce cancer rates. Or, a sociologist might be interested in spelling out the social mechanism explaining why smoking is a phenomenon increasing among women, for instance.

It is then clear that (i) not only what 'mechanism' denotes is different in the two cases (an interrelation of bio-chemical processes in the first case, and an interrelation of socio-psychological processes in the second), but also that (ii) this suggests that there exist different contexts of analysis. Some contexts need *mixed mechanisms* to be modelled. Nonetheless, this characterisation of mixed mechanisms also has some virtues. In particular, as I shall argue in the next section, it has *explanatory* virtues.

6.1.3 Explaining Through Mechanisms

As I said, causal modelling ought to be the modelling of (causal) mechanisms. One reason relates to the explanatory goal of causal models—causal models do not content themselves with merely stating descriptive claims. Of course,

explanation is not the only goal pursued by causal modelling and, as Reiss (2007) argues, modelling mechanisms might not be the most useful strategy to achieve other goals, for instance measuring concepts such as 'inflation rate' or 'unemployment rate'. But surely causation and explanation have interesting and important connections, one of them lying in the concept of mechanism. Then, the question arises as to how causal models can provide explanations. In this section I shall put forward two ideas: (i) causal models provide explanations by modelling causal mechanisms, and (ii) causal models also ought to be considered as a *model* of explanation.

To begin with, the scientific literature is not very homogenous as to the vocabulary used. *X*s and *Y*s are referred to in a variety of ways depending on the specific discipline. Statistical textbooks will normally refer to X and Y as the independent and dependent variables, or as the explanatory and response variables, respectively; the econometric literature usually talks about exogenous and endogenous variables; the epidemiological literature spell them out in terms of exposure and disease, etc. Let us focus on the explanatory-response vocabulary, which perhaps constitutes the background of all disciplines that use causal models. In this case terminology is quite explicit: the *X*s supposedly *explain* Y.

With much disappointment to the philosophers, in the scientific literature there is no explication of the terms 'explanatory', 'explanation', 'explain'. Intuitively, the *X*s explain Y in the sense that they 'account for' Y, namely the *X*s are relevant causal factors that operate in the causal mechanism, which is formalised by the structural equations and the causal graph. Needless to say, this is a very unsophisticated explication of 'explanation', yet intuitively clear. Let us leave aside, for the time being, the issue of what a good explanation is, and let us focus on what explanation in causal modelling consists of.

Causal models aim at explaining the *variability* of the dependent variable. A first suggestion might then be that the more variability we can account for, the higher the explanatory power of the causal model, and the amount of variability explained will be quantified by the coefficient of determination $r^2 \cdot r^2$, defined as the square of the coefficient of correlation r, is used to determine how well a regression fits the data. It represents the fraction of variability in Y that can be explained by the variability in the *X*s, and thus is a statistical measure that allows us to determine how certain we can be in making predictions about Y in a given model. However, this statistical answer is insufficient. This is for three reasons. The *first* is that r^2 just measures the goodness of fit, not the validity of the model, and a fortiori it does not say how well the model *explains* a given phenomenon. So r^2 gives us an idea of whether the variability in the effect is accounted for, and to what extent, by the covariates we chose to include in the model. But, and here is the *second* reason, the coefficient of determination does not give any *theoretical* motive for that. *Third*, the coefficient of determination will give us an accurate quantification of the amount of variance of Y explained by the *X*s only if the assumptions are correct. For instance, r^2 can be small either because X has only a small impact on Y (controlling for appropriate covariates) and/or because the relation between X and Y is not linear. Instead, a more satisfactory (epistemological and methodological) answer lies in the specific features of causal models. In particular, among the causal

assumptions listed in section 4.3.2, those having explanatory import are especially covariate sufficiency and no-confounding, together with background knowledge.

Together, covariate sufficiency and no-confounding convey the idea that the causal model includes all and only the factors that are necessary to explain the variability of Y. Those assumptions rely on the hypothesis of the closure of the system: causal modelling assumes, so to speak, that we can isolate a mechanism within the larger social system under consideration, and that this mechanism is not subject to external influences. Thus, we can account for Y—that is for its variability—just relying on the factors we decided to include. This is indeed a strong assumption; but this is the only way to go in order to avoid an *ad infinitum* regression hunting for more and more ancestral causes, and in order to exclude the possibility that everything influences everything else in the system, thus making it impossible to identify the causal relations to intervene upon. Covariate sufficiency and no-confounding also depend closely on which variables we choose to include in the causal model. This choice, in turn, depends on background knowledge.

From sections 6.1.1 and 6.1.2 we have an idea of what the causal mechanisms modelled in quantitative causal analysis are. But this is not enough yet to say that a causal model provides an explanation, as I haven't said anything about its formal structure. The formal structure of an explanation is given by the hypothetico-deductive methodology of causal models discussed in section 3.2, which, recall, involves three stages: (1) hypothesise, (2) build the model, and (3) draw consequences about the empirical validity or invalidity of the model. I shall not linger on the H-D methodology as it has been discussed at length in section 3.2, but I shall illustrate the hypothetico-deductive structure of causal model-explanations by means of a case study on job satisfaction next.

A Case Study: Job Satisfaction. Consider the phenomenon of job satisfaction. A study dated 1977 (Kalleberg 1977) attempted to explain job satisfaction through two variables: work values and job characteristics. Whether or not the results of this study are still valid today will not concern us—the paper still is a good example of the hypothetico-deductive character of causal models explanations. The importance of correctly explaining job satisfaction goes beyond the cognitive goal of understanding social phenomena—the explanation, as Kalleberg (1977: 124) says, also has to be useful for improving the work experiences of people. In other words, an action-oriented goal is also at the very basis of causal modelling.

To begin with, this study is not merely an empirical analysis of some data concerning job satisfaction, but it overtly aims at developing a *theory* of this phenomenon. In a footnote, Kalleberg (1977: 126 n2) points out that his paper is

> […] an attempt to develop a "theory" of job satisfaction, i.e. a set of generalisations that explains the variation in this phenomenon on the basis of the condition and processes that produce this variation […]. There are two parts to the theory: a "psychological" part which explains the variation in job satisfaction produced by the interplay between work values and job rewards and a "sociological" part, which relates the variation in job satisfaction to factors that affect one's degree of control over the attainment of job rewards.

This quote contains too much important information to be confined in a footnote. Let us disclose it. The first thing worth noting is the explicit explanatory goal

of the paper: to pinpoint the factors, the processes, and their interplay—i.e. the mechanism—that underlie job satisfaction and that bring about *variations* in it. Secondly, the rationale of causality employed is a rationale of variation, not of regularity—the interest is in whether and how variations in the explanatory variables bring about variations in job satisfaction, not whether the latter regularly follows the former. Thirdly, for the theory to be explanatory, there have to be two parts: one psychological and the other sociological, that is to say Kalleberg needs to model a *mixed mechanism*. I will now go through the paper highlighting (some of) the key elements of the hypothetico-deductive model of explanation discussed above.

The specific modelling strategy chosen here does not come out of nothing, but instead has its origins in previous works in the area. In particular, Kalleberg mentions three approaches. The first attempted to explain variation in job satisfaction merely in terms of personal traits of individuals, but this proved unsuccessful because the relation between job satisfaction and the characteristics of the job—i.e. the sociological part—was neglected. The second approach explained variations in job satisfaction just as a function of differences in the nature of jobs people perform. But this view, argues Kalleberg, had theoretical pitfalls that made it quite useless for a full understanding of the phenomenon, in spite of its practical utility. A third view was that job satisfaction has to be seen as the result of two components: the objective properties of the job and the motives of the individual. However, this approach couldn't establish those links in a systematic way, nor could it provide an adequate conceptualisation of such a mechanism.

These three approaches constitute the background from which the causal hypothesis is formulated and from which the explanatory agenda is set up (Kalleberg 1977: 126):

> The objectives of this paper are to conceptualize and empirically examine: (1) they way work values and job rewards combine to influence job satisfaction and (2) the factors that determine the extent to which individuals are able to obtain job rewards.

To establish those causal claims will require a thorough discussion of the concepts involved, of the assumptions underlying the model, of the results of empirical testing, of the theoretical plausibility of the causal links, and it will also require a comparison with alternative models.

In the paper, the discussion around the concepts involved in the causal model concerns the unitary dimension of 'job satisfaction' and the explanatory role of the variables 'work values' and 'job reward' (see Kalleberg 1977: 131). An interesting issue concerns the possible objection of a correlation between the two dimensions 'value' and 'reward': having defended the choice of the covariate and having provided arguments for their sufficiency in explaining variations in job satisfaction, the next step is to rule out a situation of confounding. I will not go through the whole of the arguments provided by Kalleberg. It will suffice to mention that he supports his claims by showing results of performed tests in various tables; in particular, covariate sufficiency is supported by showing the amount of total variation those variables are able to account for. Although statistical evidence supports the hypothesis that values and rewards are independent causes of job satisfaction, a theoretical

explanation of these results is not provided yet. In other words, Kalleberg (1977: 132) has to spell out the mechanism, that is he has to provide:

> […] a precise specification of the manner in which particular values and rewards combine to influence overall job satisfaction.

This is done by regressing job satisfaction simultaneously on values and rewards in an additive and linear model where job satisfaction is a function of rewards and values (plus errors). Testing leads to the following general result (Kalleberg 1977: 133):

> […] the highest levels of job satisfaction will be experienced by those workers with high rewards and low values, while the lowest levels of job satisfaction will be experienced with low rewards and high values.

The question arises as to whether this constitutes a good explanation of the job satisfaction phenomenon. The statistical answer is given by the good value of the coefficient of correlation obtained, and by comparing this coefficient with the one obtained in a different model, that is an alternative model having an interaction effect between values and rewards. Kalleberg argues that this second model would be less appropriate on the following grounds: (i) the first model accounts for more variance than the second, that is it has a better coefficient of determination, and (ii) empirical results of the second model go against the assumptions of the model itself.

Let me skip over the conceptualisation and modelling of the degree of control on job rewards and concentrate instead on the scope and limits of the model. Although Kalleberg makes a good case for the causal hypothesis that work values and rewards have independent effects on job satisfaction, he warns us that the 'job-reward' variable measured *perceptions* of job characteristics. The relationships between actual and perceived job characteristics are indeed an important area to investigate further in order to better understand the source of individual differences in job satisfaction. A second element worth noting is related to the assumption of closure of the system. By way of reminder, causal modelling assumes that it is possible to isolate a mechanism within a larger system, and this mechanism is not subject to external influences. Kallberg's study does not consider other external influences such as the kinds of job characteristics that produce variations in the types of job rewards he analysed. This, says he (Kallberg 1977: 140), is a key question for future research. A third consideration on future research consists in investigating what kinds of people have different values towards work, as there is a whole range of social factors that could possibly influence this variable (Kalleberg 1977: 141). Finally, a full understanding of this phenomenon needs an interdisciplinary analysis, that is it needs modelling a mixed mechanism where social and psychological factors explain job satisfaction.

Kalleberg provides a sound hypothetico-deductive explanation of job satisfaction by following the modelling procedure of causal models. In particular, we are told how the causal hypothesis is formulated, and the explanatory role of the covariates he chooses is supported by statistical and theoretical arguments, thus confirming the causal hypothesis. Also, those results are not written on stone and further research, as Kallemberg himself indicates, may complete or change the explanation of the job

satisfaction phenomenon through a new hypothetico-deductive explanatory procedure.

To sum up, a causal model attempts to *explain* a given social phenomenon—in particular, the variability of the effect variable Y—by means of a number of explanatory variables X; and the explanatory procedure is given exactly by the hypothetico-deductive methodology of causal models. How do we evaluate the goodness or the success of the explanation then? We have seen before that the coefficient of determination is insufficient to provide such an answer, which instead lies in the peculiar features of causal models. Statistical tests, notably invariance and stability tests, provide the accuracy of measurements but cannot alone guarantee the explanatory goodness of the causal hypothesis. In fact, nonsense correlations, such as the monotonic increase of bread prices in England and sea-level in Venice, may well turn out to be stable or invariant and yet not causal nor explanatory at all. The goodness of an explanation cannot be assessed on statistical grounds *alone*—the story also has to be coherent with background knowledge and previously established theories, and has to be of practical utility for intervening on the phenomenon.

Thus, the problem of the goodness of explanation is mainly a problem of internal validity, with the caveat that, among various threats, coherence with the background plays a major role. This, however, makes explanation highly context-relative simply because the causal model itself is highly context-relative. This could be seen as a virtue, rather than a drawback, as restricting the scope leads to more accurate explanations. But obviously this situation raises the problem of generalising results—that is the external validity of the causal model.

It is worth emphasising that hypothetico-deductive explanations exhibit a flexibility rarely found in other models of explanation. First, they allow a coming and going between established theories and establishing theories. Established scientific theories are (and ought to be) used to formulate the causal hypothesis and to evaluate the plausibility of results on theoretical grounds. But causal models also participate in establishing new theories by generalising results of single studies. This reflects the idea that science is far from being monolithic, discovering immutable and eternal truths. If the model fits the data, the relations are structurally stable and congruent with background knowledge, then we can say, to the best of our knowledge, that we hit upon a causal mechanism that explains a given social phenomenon. But what if one of these conditions fails? A negative result may trigger further research by improving the modelling strategies, or by collecting new data, thus leading to new discoveries that, perhaps, discard background knowledge.

The hypothetico-deductive structure of explanations also allows us to control the goodness of explanation. Notably, we can exert (i) a statistical control by measuring, with the coefficient of determination, how much variability is accounted for. We can also exert (ii) an epistemic control, by asking whether results are coherent with background knowledge. (iii) A metaphysical control is also possible, as we have to make sure that there be ontological homogeneity between the variables acting in the mechanism. If such ontological homogeneity is lacking, this triggers further research for indirect causal paths that would have been previously neglected.

6.1.4 Modelling Causal Mechanisms vs. Modelling Decision-Making Processes

But this is all about building and interpreting, that is *modelling*, causal mechanisms. How about *applying* causal mechanisms? To grasp the difference between modelling causal mechanisms and modelling decision-making processes let us consider again the two types of inference that I presented in section 5.1.

In the first type of inference, data is collected and analysed by means of statistical models. Different statistical methods, which I have presented and discussed in Chapter 3, are designed to infer causal relations from large data sets. According to scientific practice, variations will be deemed causal, roughly, if they are structurally stable. This is how, for instance, epidemiologists established the causal impact of tobacco consumption on lung cancer or how they attempt to estimate the overall contribution of genetic factors to the etiology of cardiovascular disease.

In the second type of inference, the key question is how to combine causal knowledge gathered from the population with specific knowledge about a particular individual. For instance, in the case of diagnosis the problem is how to combine population-level causal knowledge of diseases with an individual's symptoms, DNA profile and medical history, to come up with a diagnosis particularised to that individual. As for causal attribution, again we have to combine population-level causal knowledge and knowledge about a particular individual; for instance to establish whether tobacco consumption or exposure to asbestos caused cancer in a particular individual, we need to correctly apply population risks and combine them with the personal and medical history of the patient. For instance, Williams and Williamson (2006) suggest combining causal models with individual-level argumentation frameworks to tackle this problem. This second category of inference falls far beyond the scope of this work, which is rather concerned with the rationale involved in the first category.

The reason why I mention these two categories of inference is twofold. Firstly, to make a distinction between modelling causal mechanisms and modelling decision-making processes, and to disclose the different assumptions behind them. Secondly, to draw the social scientist's attention to a problem highly philosophical in character but with non-negligible consequences for methodology.

On the one hand, in modelling mechanisms we make generic causal claims about the population as a whole. The causal mechanism describes how an *average causal relation* works, that is how the causal relation is distributed or varies across the individuals of the population of reference. On the other hand, modelling a decision-making process means applying the causal mechanism to a particular individual, as is the case in diagnosis or causal attribution. It is worth noting that whilst causal attribution and diagnosis implicitly assume the correctness of the causal mechanism, causal models are precisely designed to test it.

Before dealing with the second reason mentioned above, let me emphasise that a new topic for discussion just came into range. Modelling mechanisms and modelling decision-making processes suggest a further distinction: population-level causation vs. individual-level causation. Although the topic will be analysed in detail in

section 6.2, I would like to sidetrack the reader immediately from a mistaken intuition. The distinction of different levels of causation does not necessarily commit one to a questionable position in metaphysics according to which levels are both real and distinct, and causation, whatever it is, operates differently at those levels. Nor are we committed by that distinction to a sharp dichotomy between the whole (the population, in this case) and the parts (the particular individual, in this case), or to a vocabulary more familiar to the philosopher of social science, between holism and individualism. Franck (1995) already made a significant step to overcome the dichotomy. The discussion on the levels of causation will rather follow this line: how do different levels interrelate, from the top to the bottom and from the bottom to the top? In other words, how do we go from the whole population to the individual and from the individual to the whole population, passing through intermediate levels of aggregates?

And here we come to the second reason. On the one hand, causal mechanisms, as discussed earlier in this section, are modelled by means of statistical models. In these statistical models causal relations are not sharply deterministic. They are, instead, *probabilistically* characterised. Causal models are essentially probabilistic since they include a random component. On the other hand, causal attribution and diagnosis are also made in probabilistic terms: in particular, we need some method of combining population-level probabilities and single-case probabilities. Thus both categories of causal inference make essential use of probability. Any application of probability calls for a stance concerning its interpretation and in Chapter 5 I argued in favour of the objective Bayesian interpretation. We have seen that the chosen interpretation is not neutral from a philosophical nor from a methodological viewpoint. Furthermore, as the reader will remember from the previous sections, I pointed out that mechanisms are constituted by suitable *variational* relations. It is worth noting that the variation rationale does a better job than regularity also in decision-making. In fact, regularity does not automatically make sense of the single-case process, unless we believe that single-case processes are instantiations of regular ones, which leaves unexplained cases of noninstantiation.

This distinction between modelling mechanisms and modelling decision-making processes—together with the corresponding distinction between population-level and individual level inference—has gently introduced us to the following tricky issue: the levels of causation.

6.2 Levels of Causation

As mentioned above, in social science we might be concerned with two different categories of inference: a causal relation might concern the population as a whole or a specific individual. This has led a number of philosophers to distinguish between population-level and individual-level causal claims.

We have already come across this distinction. In the discussion of probabilistic theories of causality in section 2.2, we saw that twofold causality was proposed as a solution to the problem of negative causes which arose from the squirrel

example. By way of a reminder, at the population-level squirrels are negative causes for birdies, but it might be the case that in a particular situation we feel that the squirrel actually token-caused the golf ball's falling into the hole. The best solution put forward by Sober, Eells, and others was to distinguish two levels of causation, i.e. population-level vs. individual-level (the philosophical literature also uses the synonymous expression type-level vs. token-level). This distinction is supposed to regain the squirrel *qua* positive cause at the token-level: namely, at the token-level the squirrel's kick actually increases the probability of the birdie. This distinction, at least in Eells' account (Eells 1991), presupposes that causation operates differently at the two levels, for it is only by means of a different causal mechanism that the squirrel can be a positive token cause. Hausman (1998), on the contrary, supports twofold causality without postulating different mechanisms. According to him, population-level causal claims are generalisations of individual causal claims. The two sorts of causal statements will only diverge due to random elements, or to a difference between the circumstances specified in the generalisation or in the token occurrence (see Hausman 1988: ch. 5* and ch. 9.2).

Indeed, I am also very sympathetic to twofold causality, although I do not think we need to postulate that causation works differently at the two levels. From the perspective here endorsed, this is an unnecessary and nonparsimonious metaphysical assumption.

6.2.1 Twofold Causality

6.2.1.1 Eells' Account of Token Causality

For those who still feel nostalgia for metaphysics and wish to recuperate the squirrel *qua* positive cause, I will now examine Eells' account of token causation. Although the metaphysics of causality is not our primary interest here, the discussion of Eells' account will disclose a fallacious way of reasoning about the levels, and will help us to formulate further helpful distinctions, e.g. between two different senses of 'individual-level' and between generic vs. single-case causal claims.

To begin with, the relation Eells (1991: ch. 6) wishes to analyse is: *'x's being X'* at $< t_x,\ s_x >$ caused *'y's being Y'* at $< t_y,\ s_y >$, where x takes place at time and place $< t_x,\ s_x >$, and y takes place at time and place $< t_y,\ s_y >$. That is, x and y are actual instantiations or tokens of the corresponding types X and Y, respectively.

According to Eells, there is a basic probability-increase idea that is appropriate for token causation; simply put, we must look at how the probability of the later event y actually evolves around the time of earlier and later events. By depicting the evolution of the probability value of the effect y, we actually draw the trajectory of its probability. In the case of the squirrel and the golf ball, the right trajectory is supposed to show that the probability of a hole-in-one increases after the squirrel's kick. This is supposed to regain the squirrel *qua* positive causal factor. Eells draws the trajectory shown in Fig. 6.4.

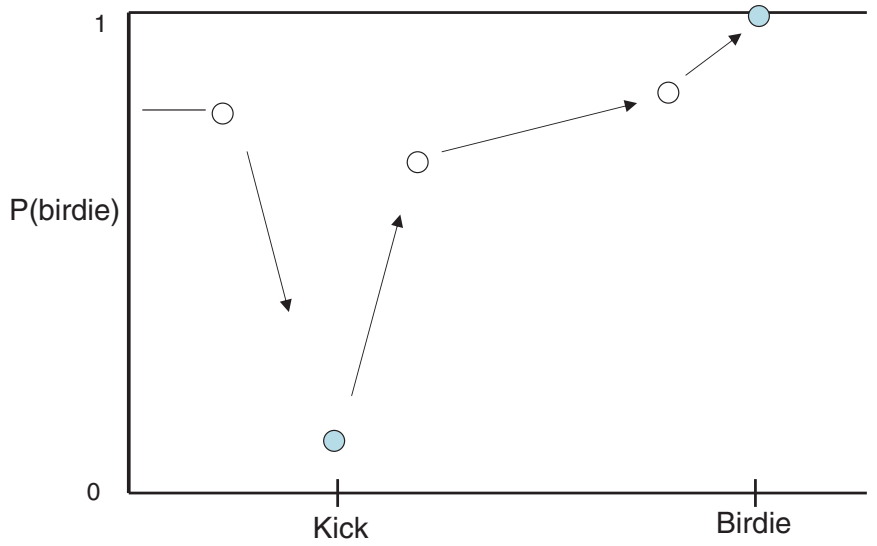

Fig. 6.4 Probability trajectory diagram for the squirrel example

As we set up the example, after the squirrel's kick at t_x, the ball is again on a path towards the cup, and this is why the ball eventually falls into it at t_y. We picture the probability of the birdie as being high just before the kick, i.e. just before t_x, then falling abruptly at the time of the kick, but then just as abruptly recovering and becoming high again, and rising till t_y, the time when the ball falls into the cup. Eells notices that, despite the fact that the probability that *the probability of a birdie* will take this trajectory is quite low—in fact the squirrel is known to be a *negative* cause at the population-level—in this particular case the probability of the birdie *does* take this trajectory. This is, in Eells' opinion, the most plausible way to enforce the intuition that the squirrel token-caused the birdie.

Let us analyse the probability trajectory in more detail. In the probability trajectory there are three crucial points in terms of which degrees of causal significance are defined.

(i) The probability of y immediately before the time t_x of the event x. This probability is the last value that $P_t(y)$ takes on before time t_x. To be more precise, this probability value would be the limit of $P_t(y)$ as t approaches t_x from the past. We assume that this limit exists and denote this value by "$\mathbf{P}^{-}$".

(ii) The probability value of y immediately after the time t_x. This probability is the first value that $P_t(y)$ takes after t_x. This probability value would be the limit of $P_t(y)$ as t approaches t_x from the future. We assume that this limit exists and denote this value by "$\mathbf{P}^{+}$".

(iii) The smallest value that $P_t(y)$ takes on the open interval (t_x, t_y). This is $\min\left\{P_t(y) : t, in(t_x, t_y)\right\}$, which intuitively is the lowest point in the region of the probability trajectory between t_x and t_y. We will denote that value by "$\mathbf{M}$".

Then, Eells says, to make sense of the squirrel as positive cause, we have to take into account its degree of causal significance. Causal significance, in Eells' theory, comes in degrees. That is, one token event can be more or less causally significant for another. For instance, an event x of type X can be a more or less strong, or a more or less weak, token-cause of a later event y of type Y. In section 6.5, Eells develops a quantitative calculus of degrees of the various kinds of token significance. The appropriate degree of significance is here "because of", that we will denote by "**B**" and define as follows:

$$B = M - P^{-}. \tag{6.1}$$

According to Eells, **B** can take values between -1 and $+1$.

Notice that **M** is always less than or equal to $\mathbf{P}^{+}$, i.e. $\mathbf{M} \leq \mathbf{P}^{+}$. This is because **M** is the smallest value of $P_t(y)$ after t_x and before t_y; since $\mathbf{P}^{+}$ is the value of $P_t(y)$ after t_x and before t_y, $\mathbf{P}^{+}$ cannot be less than **M**. The relation $\mathbf{M} \leq \mathbf{P}^{+}$ is the only constraint on the three crucial points of $P_t(y)$ mentioned above. Furthermore, it is somehow natural to assume that they are all strictly between 0 and 1, since $P_t(y)$ is 1 at t_y (so that $P_t(y)$ should never be 0) and since it is only at t_y that *'y's being Y'* is certain (so that $P_t(y)$ should not be 1 at any of the crucial points, which are all before t_y).

The degree of causal significance 'because' may also be characterised as follows:

(i) $P_t(y)$ changes at t_x.
(ii) This probability is high just before t_x.
(iii) This probability is higher right before t_x than after t_x.
(iv) This probability remains high until t_y, at which point it becomes 1.
(v) The degree of 'because' is determined by how much the probability of y changes at t_x, by how high this probability is just after t_x, by how much higher it is at later times in the interval (t_x, t_y), and by how high it remains after t_x, until it becomes 1 at t_y.

The value $\mathbf{B} = \mathbf{M} - \mathbf{P}^{-}$ is supposed to be an appropriate measure of the degree to which *'x's being X'* succeeds in increasing the probability of *'y's being Y'*. If **B** is positive, then the probability of *'y's being Y'* is always higher after t_x, than it was before t_x. Thus, positive values of **B** mean that *'y's being Y'* occurred because of *'x's being X'*. When **B** is negative or 0, then *'y's being Y'* was not at all because of *'x's being X'*.

So far so good. This is an extremely complicated way of saying that positive token causes raise the probability of their effects. Eells' probability trajectory might match our intuitions about the token squirrel. However, in order to say that the squirrel token-caused the birdie, we have to calculate probabilities and show that in this particular case there has been a probability increase after the kick.

Here is Eells' solution (Eells 1991: 344 ff.). For all times t:

$$P_t(y) = P(y|K_a \cap W_t). \tag{6.2}$$

where K_a represents the causal background context obtained by holding fixed all factors of the following two kinds:

(i) Any factor Z such that:

(a) Z is exemplified in the case in question.
(b) Z's exemplification is token uncaused by *'x's being X'*.
(c) Z is type-level causally relevant to *'y's being Y'* in the context determined by the way things are before t_z.

(ii) Any factor Z such that:

(a) Z is exemplified in the case in question.
(b) Z's exemplification is token uncaused by *'x's being X'*.
(c) X interacts with Z with respect to Y in the context determined by the way things are before t_z.

W_t is instead the conjunction of all factors $F_{t'}$ such that:

(a) $F_{t'}$ is not included in K_a.
(b) $F_{t'}$ has fallen into place at or before time t.
(c) $F_{t'}$ is type-level causally relevant to y in the context K_a,

and with the stipulation that x be included in any W_t for the same time t_x or later, and y be included in any W_t for the same times t_y or later.

I would not blame the reader who got lost in this endless list of obscure factors, for I did too. This is the clear signal of troubles. A first problem concerns the exact specification of K_a and W_t. According to Eells, the probability trajectory of the golf ball can be evaluated in the crucial points, provided that we have at our disposal *complete* and *exact* knowledge of the causal background K_a and of the conjunction of the causal factors in W_t. Unfortunately, such knowledge is very seldom available. To say that *in principle* the state of the world would allow an extremely precise calculation of the probability of the birdie does not mean that *de facto* we are able to calculate it. At best we can compute expected values on the basis of available knowledge, by updating probabilities. Therefore, we should do with the little we have, and according to this little available knowledge, the squirrel, as far as we know, *is not* a positive cause. This is what epistemology says.

A second difficulty concerns the interpretation of probability here adopted. No doubt the choice of an objective and physical interpretation explains why Eells tries to spell out token causation in terms of probability trajectories. Namely, the trajectory is supposed to trace the *physical evolution* of the probability of 'y's being Y' from time t_x until the realisation at t_y, where y actually occurs and hence gets probability 1.

Let me emphasize that this line of argument is ambiguous, for *post hoc* probabilities are always 0 or 1. So, at the time t_y, of course $P(y)$ will be either 0 or 1. Eells' probability trajectory is somehow misleading. His line of reasoning seems to be: once the birdie occurs at t_y, $P_t(y) = 1$, so, let us try to draw the best probability trajectory that makes sense of the following: (i) the *post hoc* probability is 1, and (ii) we strongly believe that the squirrel token-caused the birdie. In other words, to draw such a probability trajectory is our best guess to explain the hole-in-one—it is a probability trajectory drawn *ad hoc*—but Eells fails to tell us *what* epistemic

access we do have in order to draw such a trajectory. The sole epistemic access he calls into his argument is our *confidence*, being witnesses to the shot, that the squirrel caused the birdie. This is *epistemologically* weak.

A third difficulty concerns the evaluation of probabilities at the crucial points along the trajectory. I understand Eells' view that these probabilities refer to the hypothetical population the token factor under consideration belongs to, but indeed this is the problem. Let me explain this passage. As Eells (1991: 44 ff.) puts it, we begin with an actual population P, and we are interested in conditional and unconditional probabilities involving certain factors, say X, Y and Z. Let us think of the actual population P as the result of an 'experiment', i.e. an experiment that could, theoretically, be repeated.

As a result of the experiment, we now have at our disposal conditional and unconditional frequencies involving the factors X, Y, Z, etc. in the population P. In this experiment, initial conditions were distributed among the N members of the actual population P in one certain way. So, when we repeat the experiment we should distribute the very same conditions in the very same way, although the population will be a different one. We assume that the same set up could be done again and again, and we expect the resulting frequencies to differ from one experiment to another, exactly because distinct experiments are run over different populations. Hence, according to Eells, if we wish to evaluate $P_t(y)$ at different times along the trajectory, we have to consider different hypothetical probability distributions that hold respectively at t_x, or at any time t between t_x and t_y, or at t_y.

Although I do think that probabilities in the single case *are* evaluated thanks to knowledge of the corresponding population, my doubts here concern (i) the very possibility of being able to evaluate at *every* time the corresponding hypothetical distributions and, (ii) the fact that if at different times we evaluate probabilities in terms of different hypothetical distributions, we miss the connection to the actual population the token under analysis belongs to.

6.2.1.2 Epistemological and Methodological Morals

Before giving a more sensible interpretation of the levels of causation, let me clarify one issue. So far, we have been using the terms population vs. individual-level and type vs. token-level indistinguishably. Although in the philosophical literature those pairs of oppositions are used interchangeably, this is the source of misunderstandings with practising scientists. The source of confusion is dissolved once we use the following distinction instead: generic vs. single-case. If the reader now fears that she will soon get lost, I beg her to be patient because in a few paragraphs the issue will be clarified also from a methodological viewpoint.

To begin with, the problem raised by philosophers is not trivial at all; however, it does not quite match with the distinction population vs. individual as scientists draw it. The source of confusion lies in the fact that causal models can handle, as we shall see later, aggregate or individual variables. Thus scientists tend to identify this distinction with the philosophers' one. However, the discussion of modelling

mechanisms vs. modelling decision-making processes should have already shed some light on the issue. Whether they use aggregate or individual variables, causal models aim at establishing *generic* causal claims and in this sense both correspond to the philosophers' population- or type-level. On the other hand, we might be interested in a particular instantiated causal relation, and in this case it would be better talking about *single-case*. This corresponds to the philosophers' individual or token-level, but has no corresponding term in the scientific jargon.

In sum, if we do not want to appeal to the assumption of different causal mechanisms, what meaningful understanding of single-case causal relations can be given? The adoption of a probabilistic perspective should make the squirrel harmless for epistemology and methodology. In fact, a probabilistic approach includes from the start the possibility of non-instantiated cases or of outlying observations. The squirrel example (or the case of Harry being a heavy smoker and yet never developing cancer) raises a problem for the metaphysics of causality: is causation (whatever *metaphysically* it is) the same at all levels of reality? From the epistemological and methodological standpoint I endorse, I do not need to develop a comprehensive account of single-case causation. I suggest the following *epistemological* understanding, though.

Single-case causal relationships can be thought of as realisations of joint probability distributions. From a mere statistical viewpoint, at the generic-level, causal relations are represented by joint probability distributions, the single-cases are realisations thereof. This gives a sensible answer to the squirrel example. The information encoded in the probability distribution at the generic level already includes the possibility of a joint realisation of a 'hole-in-one' and a 'squirrel kick'. This probability exists, although it is very tiny. If this probability is very small, this observation lies in the queue of the distribution, as statisticians say, but it definitely can happen. Also, if we randomly pick up a squirrel from the population, there is no reason not to believe that it will follow the generic trajectory—i.e. it will be a preventative for birdies. This causal belief will be questioned only *post hoc*, in case the squirrel kicks the ball and makes it fall in one.

The same line of reasoning holds if we want to provide an understanding of non-instantiated cases or outlying observations. To think of single-case causal relations as realisations of joint probability functions *epistemologically* makes sense of how it can be the case that the golf ball falls into the cup, despite the squirrel's kick. More importantly, it makes sense of how it might be (statistically) the case that Harry did not develop lung cancer, in spite of his heavy smoking.

Before closing this section, let me emphasise two important issues about twofold causality. *First*, this understanding of twofold causality is metaphysically parsimonious, since it abstains from postulating different causal mechanisms at the generic level and at the single-case level. *Second*, the distinction between the epistemology of causality and the metaphysics of causality is fundamental. In fact, *single-case* causation clearly is *ontologically* primitive, since populations are made up of individuals among which single-case causes and not generic causes operate. Nonetheless, *generic* causation is *epistemologically* and *methodologically* primitive. This is for two reasons. First, causal conclusions drawn from statistical analyses

concern the population, not directly real individuals in the population, and second, to assign probability values to individuals, e.g. in the case of diagnosis or causal attribution, we refer to the corresponding generic probability values. However, this still leaves unanswered the question of how to reason about single-case causal relations relying on generic causal knowledge, which will be the object of the next section.

6.2.2 The Connecting Principle

It is a problem of the epistemology of causality to tell us how these levels of causation can be connected. For this purpose let us see whether the unduly neglected Connecting Principle can be useful. Stated in Sober's 1986 paper, and almost entirely neglected both in the philosophical and scientific literature, the Connecting Principle (henceforth CP) has nevertheless great intuitive appeal.

The problem to be solved can be put as follows. Causal conclusions drawn from statistical analyses concern the population, i.e. they are generic causal claims, but we might be interested in single-case causal claims as well. These are two different issues. The first concerns the way priors are assigned to single-cases, while the second is concerned with the likelihood of hypotheses. In truth, CP deals directly with the second problem; the first problem, as I shall show, will find its solution throughout the discussion. In the following, to be faithful to Sober's exposition, I will be using the terms type/token instead of generic/single-case. Here is how Sober formulates the idea behind CP (Sober 1986: 101):

> If a token event of type C is followed by a token event of type E, then the support of the hypothesis that the first event token-caused the second increases as the strength of the property causal relation of C to E does.

Thus Sober's formalisation of CP (ibidem):

> If C is a causal factor for producing E in a population P of magnitude m, then S {C (t_1) *token caused* E (t_2)|C (t_1) *and* E (t_2) *occurred in the population* P} = m.

Here C and E represent properties, and $C(t_1)$, $E(t_2)$ denote events, i.e. the events occurring at space-time t_i and having the property C and E, respectively. In his formulation, Sober makes use of $S(H|E)$ that expresses the support of the hypothesis H given the evidence E, and that can take values between -1 and $+1$.

It seems to me that there is a slight discrepancy between the rough idea and the formulation of CP. Whilst the first is intuitively right, the second is a bit imprecise. Consider Sober's notation: C, E represent properties or types, and $C(t_1)$, $E(t_2)$ denote occurred events. Let us now denote type-level causal relations by $C \rightarrow E$, and token-level causal relations by $C(t_1) \rightarrow E(t_2)$. For brevity, I shall call *token hypothesis* the hypothesis that $C(t_1)$ token-caused $E(t_2)$, and *type hypothesis* the hypothesis that C causes E in a given population P. Roughly, Sober says that the support of the token hypothesis is proportional to the strength of the causal relation at the type level, namely the *support* of $C(t_1) \rightarrow E(t_2)$ is *proportional* to the *strength* of $C \rightarrow E$. The formulation states something different, however. It states that if the magnitude

(i.e. the *strength*) of $C \rightarrow E$ is m, then the support of $C(t_1) \rightarrow E(t_2)$, given that $C(t_1)$ and $E(t_2)$ occurred, will be m as well.

Additionally, according to the formulation of CP, Sober is not interested in the support of the token hypothesis given that the relata *occurred*, as the formalisation states. Rather, Sober seems to be interested in the support of the token hypothesis, given that at the population level the corresponding causal relation holds and has strength m. These remarks motivate a reformulation of CP, in order to make the formalisation consistent with the intuition behind it. More specifically, I will try to reformulate CP exploiting $S(H|E)$, the support of a hypothesis given certain evidence. So the first step will be to make explicit what H and E in CP are.

Let us begin with E. E is our evidence, or prior knowledge when we are interested in the assessment of support for token hypotheses. Thus, in accord to what is stated in the rough idea, E denotes the type-level causal relation, having a certain strength m. H, instead, refers to the token-level causal relation; in particular, H refers to the token hypothesis under analysis. Furthermore, lower case letters *c, e* will refer respectively to causes and effects at the type-level, and c_t and $e_{t'}$ will denote the corresponding tokens. As before, m will be the magnitude or strength of the type-level causal relation.

Thus CP*:

> If c is a causal factor of magnitude m for producing e in a population P, then $S(H|E)$ is proportional to m.

CP*, thus reformulated, says that if we know that a certain type-level causal relation has strength m, then the support of the corresponding token hypothesis, given this evidence, is proportional to m.

Although this reformulation seems now consistent with the first formulation, in order to understand the compass of CP* we need to discuss three issues. (i) How is the magnitude m of the population-level causal relation computed? (ii) What exactly is $S(H|E)$? And (iii) why has the equality been replaced by proportionality?

Strength of the Causal Relation. To assess the causal strength Sober proposes the following strategy. We compute the average impact of a causal factor in the population, taking into account other factors in the background context. Thus the proposal (Sober (1986: 102, my notation). Lower case letters denote factors at the population-level, following the new notation for CP*):

> If c is a causal factor for producing e in a population P, then the magnitude m of the causal factor is:
>
> $$\sum_{i=1} \{P(e|c \cap b_i) - P(e|\overline{c} \cap b_i\} \cdot P(b_i).$$

In order to give an intuitive idea of how this works, let us think of the impact of smoking (c) on coronary heart disease (e). To assess the causal strength of smoking, we compare what happens when the factor is present to what happens when the factor is absent. Presumably, the probability of coronary heart disease is higher when the factor smoking is present than when it is absent, *given* the same specified set

of background conditions (i.e. the b_is). By this strategy, one measures the average difference the causal factor makes in the conditional probability of the effect. The causal strength thus computed will fall between 0 and 1 when the causal factor is positive, and between −1 and 0 when the causal factor is negative.

The choice of the strategy to assess the causal strength is not of secondary importance; however, this is not my main concern at the moment. Others would prefer adopting a different measure of correlation. Besides, the debate on the adequacy of correlation coefficients for measuring the causal strength is still open, and I am not going to bring it to an end here. This is for two reasons. The first is that the point at stake is not the adequacy of one or another method to assess the causal strength, and the second is that, as Ellett and Ericson clearly pointed out, *different* measures of correlation will give consistent qualitative results, i.e. those results will all agree on whether the factor is positive or negative (see Ellett and Ericson 1983, 1984, 1986a, b). Precisely, Ellett and Ericson argue that for dichotomous variables different measures of correlations will give a coherent *qualitative* measure of the causal impact, but not a coherent *quantitative* one. For probabilistic causal models dealing with dichotomous variables they propose an approach based on conditional probabilities (Ellet and Ericson 1984). So, as long as different strategies are coherent, i.e. they detect the causal strength in the same direction (either positive or negative), for the present discussion it does not matter which one is preferred.

Likelihood and Support. Without going through the entire history of statistics, to grasp the intuitions behind the concept of support we need to take one step back to the Fisherian concept of likelihood. Fisher (1922) used likelihood as a predicate of hypotheses in the light of data. It is worth noting that 'likelihood' can be also used as a predicate of data, in the light of hypotheses. As we shall see later, this second use is exactly the 'likelihood' occurring in Bayes' formula. So, the likelihood of a hypothesis relative to an observation indicates how well the observation supports the hypothesis.

Edwards (1972: 9–10). defines the likelihood $L(H|E)$ of the hypothesis H given evidence E as being proportional to $P(E|H)$, and the likelihood ratio of two hypotheses on some data as being the ratio of their likelihoods on the same evidence:

$$\frac{L(H|E)}{L(\overline{H}|E)} = \frac{P(E|H)}{P(E|\overline{H})}. \tag{6.3}$$

It is worth noting that when we handle probabilities, E is variable and H is constant; whereas when we handle likelihoods H is variable and E is constant. There are other differences between likelihoods and probabilities, the most important of which is that probabilities, according to Kolmogorov's axiomatization, are additive and likelihoods are not. I will not go through the technicalities of likelihoods, for which I direct the reader to the works of Edwards (1972), Hacking (1965), and, of course, Fisher (1922, 1930, 1935).

Rephrased for our current problem, $L(H|E)$ is the likelihood of the hypothesis that c_t token-caused $e_{t'}$, given that the corresponding type-level causal relation holds and has a certain strength m. This likelihood is proportional to the causal strength,

namely the stronger the type-level causal relation, the more likely the hypothesis about token-level causal relation.

Let us examine the concept of likelihood in more detail. Relying on the Fisherian concept of likelihood, Edwards defines the support S as being the natural logarithm of the likelihood ratio $L(H|E)/L(\overline{H}|E)$. It is easier to grasp the intuition behind the use of the likelihood ratio rather than its natural logarithm. The likelihood ratio is the ratio between the frequencies with which, in the long run, either H or *not-H* will deliver the observed data. Log-likelihoods are preferred because they satisfy several *desiderata*, e.g. transitivity, additivity, invariance under transformation of data and invariance under transformation of parameters. For the technical details see Edwards (1972: 28–30).

One nicety of likelihoods is that they allow the comparison among different sets of data or different hypotheses. This means that we can measure the support of the hypothesis that Harry developed lung cancer because of his heavy smoking, given that in the population smoking is a considerable risk factor. But we could also measure the support that it was asbestos that caused him to have lung cancer, given the same results at the population-level. This can be done by working out the likelihood ratio $L(H_1|E)/L(H_2|E)$, where H_1 denotes the hypothesis that asbestos caused cancer, and H_2 the hypothesis that, instead, smoking caused cancer. How likely the hypothesis about asbestos is depends on available knowledge about the strength of different causal factors in the population of interest. It is worth stressing that CP* does not intend to assign a magnitude to the token causal relation. CP* just says how likely the token hypothesis is.

Proportionality Instead of Equality. The choice to replace equality by proportionality in CP* better mirrors the intuition expressed in the rough idea: the higher the strength of the type-level relation, the greater the support of the token relation. So, CP* says that for greater values of m, the support of the token hypothesis will be greater as well. Besides, there are two problems if we equate the support $S(H|E)$ with the measure of strength m.

In the first place, I do not see why the support S, according to Sober, should vary between -1 and $+1$. As I just mentioned, the support is defined as the natural logarithm of the likelihood ratio, and natural logarithms vary between $-\infty$ and $+\infty$. On the other hand, the strength can vary depending on the measure of correlation adopted, as Ellett and Ericson proved, but in any case m falls in the interval $[-1;1]$; consequently, it seems odd to state that the value of S should be exactly equal to the value of the magnitude.

Notice, however, that CP* does not say how to assign prior probabilities in the case of token hypotheses, but it seems reasonable, in the absence of additional knowledge, to plug in the type-level results, as is customary in computation of risk factors.

The last remark concerns the concept of likelihood and the interpretation of probability. The Fisherian concept of likelihood is intended as a measure of belief. Consequently, this subjective understanding of likelihoods fits the objective Bayesian interpretation of probability preferred for the single-case. In other words, this use of

the support $S(H|E)$ is not in contraposition to Baysianism. In fact, likelihoods enter Bayes' theorem as follows. Remember that Bayes' theorem

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \tag{6.4}$$

can be rewritten in terms of odds and likelihood ratios:

$$\frac{P(H|E)}{P(\overline{H}|E)} = \frac{P(E|H) \cdot P(H)}{P(E)} \cdot \frac{P(E)}{P(E|\overline{H}) \cdot P(\overline{H})} = \frac{P(E|H)}{P(E|\overline{H})} \cdot \frac{P(H)}{P(\overline{H})}, \tag{6.5}$$

and here $P(E|H)/P(E|\overline{H})$ is exactly the likelihood ratio $L(H|E)/L(\overline{H}|E)$ for the two hypotheses H and *not-H* on evidence E as defined above (see also Edwards 1972: ch. 4).

Let me summarise. Sober interprets the support $S(H|E)$ as the (natural logarithm of the) likelihood ratio of the hypothesis relative to the available evidence. This approach is consistent with the intuitions about the two levels of causation and about the corresponding interpretations of probability. In fact, the strength of a causal relation at the population level is assessed by the computation of a measure of correlation. This computation is based on the data concerning the population. Subsequently, knowledge concerning the causal relation at the population-level is used to allocate priors in the token case and to establish the support of the corresponding token hypothesis.

In the end, CP* allows us to relate the two levels of causation *epistemologically*: from prior knowledge about the population we are able to assess the support of the corresponding token hypothesis. Differently put, CP* helps in *epistemologically* relating population-level causal claims and individual-level causal claims. That is to say, CP* is not used to assign the strength of the causal relation; rather, CP* tells us what is to be expected at the token level. Moreover, CP* mirrors the idea that generic causal relations are epistemologically primitive, while single-case causal relations are ontologically primitive, that is, how claims about generic causal relations bear on facts about single-case causation.

6.3 Levels of Analysis

Although the generic vs. single-case distinction has proved to be very useful, this section goes further in arguing that the problem of the levels of causation is better understood and dealt with if reformulated as a problem of *levels of analysis*. In turn, dealing with the levels of analysis requires a prior clarification of the *elements* involved, viz. the types of variables, and then how those elements articulate in different types of analysis. I shall then discuss the extent to which multilevel analysis sheds light on the interrelations between different types of variables across levels of analysis and across disciplines. Finally, the rationale of variation will be shown to have a great deal on these issues.

6.3.1 Types of Variables and of Fallacies

6.3.1.1 Types of Variables

Practising scientists will be surely familiar with the various sorts of variables they handle. An overview is nonetheless useful for the discussion about on the types of analyses that will follow. A taxonomy can be offered according to these criteria:

- Genre and scale: continuous/discrete, quantitative/qualitative.
- Role: explanatory/response, observed/latent/instrumental/proxy.
- Level: individual/aggregate.
- Field: socio-economic/demographic/biological/....

Genre and scale. A *discrete* variable is one that may take only a countable number of discrete values such as 0, 1, 2, 3, etc. Examples of discrete variables include the number of children in a family, the Friday night attendance at the cinema, the number of patients in a doctor's surgery, the number of defective bulbs in a box of ten. A *continuous* variable is one which takes values in an interval of the real numbers. Continuous variables are normally measurements. Examples include height, weight, the amount of sugar in an orange, the time required to run a mile. A *quantitative* variable is measured on a numeric or quantitative scale. Examples include a country's population, a person's shoe size, or a car's speed. A *qualitative* variable refers to information grasped through classes. Qualitative variables are measured according to different levels of measurement. A *categorical* or nominal scale means that there is no intrinsic ordering among the categories; examples include religion (Catholic, Muslim, Jewish, ...), marital status (single, married, divorced, widowed, ...), and gender (male, female). An *ordinal* scale means that there is a clear ordering, for instance economic status may be classified as low, medium or high. An *interval* scale means that values are ordered in equal intervals; for instance, in the Fahrenheit scale of temperature, the difference between 30 degrees and 40 degrees represents the same temperature difference as the difference between 80 degrees and 90 degrees.

Role. Variables are called *explanatory* or *response*, according to the (causal) role they play. The response variable, also called 'dependent', is the effect we are interested in, and the explanatory variable(s), also called independent, is (are) the putative cause(s). Variables can be *observed*, and thus directly measured, or unobserved, in which case are also called *latent*. A latent variable is a variable that cannot be measured directly but is supposed to underlie the observed ones. Examples include quality of life, self-confidence, morale, happiness, etc. A *proxy* is then an indicator describing a variable that is not directly observable; for instance, as student motivation is not directly observable, class attendance might be used as a proxy. *Instrumental* variables are used to produce a consistent estimator when the explanatory variables are correlated with the error term. An instrumental variable has to be itself correlated with the explanatory variable but it cannot be itself correlated with

the error term and it has to bear an effect on the response variable only through the explanatory variable, not directly.

Level. An *individual* variable measures individual characteristics, i.e. it takes values of each of the lower units in the sample. For instance, we can record the income of each individual in the sample. An *aggregate* variable for a higher level unit (e.g. a family, a class, a firm, …) is constructed by combining information from the lower level units of which the higher level is composed (e.g. family members, students, employees, …). Thus an aggregate variable is a summary of the characteristics of individuals composing the group—for instance, the percentage of farmers in the migration example (Chapter 1) or the mean income of state residents.

Field. Variables can store information about individual or group characteristics that pertain to specific fields or domains. A demographic variable is a characteristic representing a social or vital statistics of an individual, sample group or population; for instance: age, sex, racial origins, marital status. A socio-economic variable describes socio-economic characteristics as for instance level of education, unemployment rate, economic status indicators, etc. Other variables may pick up on aspects more related to the socio-*political* characteristics of the population, for instance political ideology and behaviour, attitudes towards salient public policies. Other variables may describe the *psychological* characteristics of individuals such as self-esteem, shyness, IQ, etc. Variables can also store *biological, medical, health* information about individuals. All these types of variables are not rigidly fettered in their respective domains but, as shown in mixed mechanisms, they can and ought to be used across fields of research.

6.3.1.2 Atomistic and Ecological Fallacies

The atomistic and ecological fallacies are quite well known to social scientists. As I mentioned earlier in section 3.1.5, both fallacies concern inferences. The atomistic fallacy occurs when we wrongly infer a relation between units at a higher level of analysis from units at a lower level of analysis. The ecological fallacy goes the other way round: it draws inferences about relations between individual-level variables based on the group-level data. The first person to draw attention to this problem—the one who coined the term 'ecological fallacy'—was William Robinson in his 1950 paper "Ecological correlations and the behaviour of individuals". He analysed each of the 48 states in the US with the 1930 census data. He computed the literacy rate and the proportion of the population born outside US. Robinson firstly found a positive correlation of 0.53 between the two variables, i.e. the higher the proportion of immigrants, the higher the literacy rate. However, after a finer examination of individual-level data, the correlation appeared to be negative (−0.11), i.e. immigrants were on average less literate than citizens. The misleading positive correlation between aggregate quantities was due to the fact that immigrants tended to settle down in states where the native population was more literate. In this case aggregate variables did not reflect individual behaviour.

6.3.2 Levels of Analysis vs. Levels of Causation

Once types of variables and of fallacies are taken into account, the problem of the levels of causation becomes a genuine methodological problem with significant philosophical consequences. Let us see why. Let me emphasise once again that the problem of the levels of causation as philosophers have addressed it and as practising scientists understand it are quite different things. On the one hand, philosophers seem to assume that causal laws (or weaker causal generalisations) can be established and they wonder how some theories—particularly, probabilistic theories—deal with cases of noninstantiation or of outlying observations. In other words, given a causal generalisation such as the trivial one 'squirrels' kicks lower the probability of birdies' or the more serious one 'smoking raises the probability of developing lung cancer', they wonder whether a probabilistic theory of causality is still coherent when it faces observations that do not instantiate the causal generalisation. On the other hand, practising scientists rather reason with variables and with probability distributions. Thus, to them the population vs. individual is understood as aggregate vs. individual. That is to say, the population vs. individual distinction is a distinction that concerns the types of variables (aggregate or individual) and the types of model (aggregate or individual). However, the philosophers' problem, although not overtly addressed in the scientific literature, seems to be a concern for practising scientists. Witness for instance Parascandola and Weed (2001: 906):

> Another concern [about a probabilistic definition of causality] is that the definition remains unclear about what it means to say that smoking *raises* one's probability of developing cancer. (Emphasis in the original.)

This is indeed a restatement of the philosophers' problem: given a causal generalisation, what does it actually mean in the single case? Similarly, Lagiou et al. (2005: 565) defend a probabilistic definition of cause and then wonder what the implications are for a particular individual:

> In terms of a particular individual, exposure to a cause of a certain disease implies that the individual is now more likely to develop the disease, although there is no certainty that this will happen.

Inference of causation in a particular individual is not straightforward, though, and in fact Lagiou et al. distinguish three levels of analysis. *Level I* corresponds to a particular study but, needless to say, causation can never be inferred from a single one. *Level II* is the general case, where, on the basis of several particular studies and of the evaluation of a number of other criteria, we establish the aetiological role of a particular exposure. *Level III* corresponds to disease in a particular person, i.e. it deals with the possibility of accurate causal attribution on the basis of results in Level II.

Granted, not all social sciences are concerned with the single case, i.e. real and concrete individuals, to the same extent. Many disciplines are more concerned with the generic level and only indirectly with the single-case or the other way around. Demography, for instance, studies migration behaviour of populations but it is not

directly interested in the probability that a particular individual, say Harry, will migrate. On the other hand, it is often said that in medicine there are not illnesses but only ill persons to cure. Physicians, then, are more concerned with the single case. Though not directly concerned with individuals, social sciences such as sociology, demography or economics do have a bearing on the individual since their results orient and guide public policies, for instance to reduce unemployment or to discourage tobacco consumption. On the other hand, to correctly assign single-case probabilities, physicians do need to take generic epidemiological results into account.

Let us now turn our attention to the scientists' problem, namely the use of individual and aggregate variables to build individual and aggregate models.

6.3.3 Levels of Analysis

Before discussing the levels of analysis by aggregation, let me draw the reader's attention to another type of level of analysis, namely by *discipline*. That there be different levels of analysis by discipline has already emerged in the discussion of mixed mechanisms, where I argued that to give more complete and accurate causal stories we need to model mechanisms that have different components, for instance both biological and socio-economic variables. I wish to stress further this point here because the search of causes of effects may lead to different results depending on the discipline. Vineis (2003: 84), for instance, is worried that:

> Molecular biologists tend to consider relevant changes at the molecular level as the 'cause' of cancer, while epidemiologists usually refer to external agents.

He then tries to reconcile molecular biologists and epidemiologists by invoking Aristotle's four causes: molecular changes would be the *formal* cause and external agents the *efficient* cause. No doubt it is an interesting step to revive Aristotle's philosophy in contemporary scientific practice, but I shall leave to historians of philosophical thought to judge the plausibility of this application of Aristotle. Instead, in the light of the previous discussion of mechanisms, we can claim, with no contradiction, that molecular biologists and epidemiologists are modelling different mechanisms, thus belonging to different levels of analysis, viz. by discipline.

We come now to consider levels of analysis *by aggregation*. As I have already mentioned in the presentation of multilevel analysis in Chapter 3, the market offers individual models, aggregate models and mixed models, i.e. models that integrate in a single framework individual and aggregate variables. Having already presented the formal structure of these mixed models, also called multilevel or hierarchical models, we can now spend time on some issues that motivate the development of this methodology. Let us follow up Courgeau's arguments (Courgeau 2003). In particular, I would like to address three issues: (i) the non a priori specification of the level of analysis, (ii) the attempt to go beyond a holistic and individualistic approach, (iii) the concept of statistical individual.

Courgeau is surely right in claiming that there is no a priori specification of the level of analysis. The social sciences start with observing human phenomena and

behaviours and then try to structure them to give a sensible causal story or explanation. *A priori* there is no reason why demographic analyses should be run at the level of individuals, family, local population or national population; or economic analyses to be run at the level of individuals, or market or nation. If the reason cannot be specified a priori, it must lie somewhere else. Because individual and aggregate variables store different sorts of information, individual and aggregate models will potentially lead to different results. That is exactly the lesson of the atomistic and ecological fallacies.

In claming that the social sciences try to give sensible explanations of observed human phenomena and behaviours, Courgeau explicitly says that the problem of a specific individual behaviour, i.e. of the single-case, is set aside. No doubt the primary interest of multilevel models is *not* the particular individual; however, the generic vs. single-case distinction cannot be dismissed so quickly. The social sciences are in general not directly concerned with the single-case but (i) generic causal claims—whether at the individual or aggregate level—are inferences based on a number of observed single-cases, and (ii) generic causal claims and explanations must have a bearing on the single-case if they are to be effectively applied to set policies and interventions to change the behaviours of *single-case* individuals.

Let me now linger on the idea of giving 'sensible explanations'. The question that easily arises is the following: what makes an explanation *sensible*? Bunge (2004) argued that the only way to explain a given phenomenon is to show the *mechanism* that brings it about. The trouble with Bunge's view is that he holds quite a narrow view on mechanisms as they are essentially physical. An option is still to take up Bunge's intuition—that sensible causal explanations involve mechanisms—but we have to broaden our view. I have discussed at length a broader conception of mechanism that relies on causal *modelling*, which, in turn, involves several crucial elements, as for instance statistical and causal assumptions, the use of background knowledge and, in multilevel modelling, interrelations between and among different levels of aggregation.

Multilevel analysis is an attempt to go beyond holism and individualism. In the general introduction, Courgeau (2003: 3 ff.) gives us a beautiful discussion of holism and individualism in the social sciences and of the need for multilevel modelling to overcome the dichotomy. Briefly put, according to holistic approaches the properties of a given system cannot be reduced to the mere sum of its components—the system as a whole determines in a fundamental way how the parts behave. According to individualistic approaches, social phenomena and behaviours can be explained by appealing to individual decisions and actions, without invoking any factor transcending them. Holism and individualism would easily lead us into the intricate field of social ontology. Unfortunately, this falls far beyond the scope of the present work. I recommend the reading of van Bouwel (2004, and references therein) who urges us to redirect the holism/individualism on a methodological track rather than a metaphysical one. The reason why I nonetheless mention this issue is to support the empiricist stance that has accompanied the whole of this book from the very start. Our causal relata in causal modelling are *variables*—be they individual, aggregate, social, demographic or biological—and there is no need to *reify* them into

supervenient properties of populations. This is mirrored in the statistical understanding of the levels of causation mentioned before, i.e. generic causal statements can be thought of as joint probability distributions the single-case are just realisations thereof. But this is also mirrored, and even supported, in the choice of the interpretation of probability advanced in Chapter 5. The objective Bayesian interpretation allows a sensible interpretation of probabilistic statements both at the generic and the single-case level. In particular, if a frequentist interpretation is adopted at the generic level, the danger of reifying variable into supervenient populational properties is avoided.

Let us now turn our attention to the concept of *statistical individual* as opposed to a *real* individual, which is the source of misunderstandings between practising scientists and philosophers. Courgeau (2003: 8) makes the point very clearly:

> In the search for individual random processes, two *individuals observed* by the survey, possessing identical characteristics, have no reason to follow the same process. By contrast, in the search for a process underlying the population, two *statistical individuals*—seen as units of a repeated random draw, subject to the same selection conditions and exhibiting the same characteristics—automatically obey the same process. (Emphases in the original.)

A couple of thoughts to follow up this claim. Firstly, the distinction drawn before between modelling causal mechanisms and modelling decision-making processes goes precisely along these lines. Because two *observed* individuals have no reason to follow the same process, in causal attribution or diagnosis we have to combine generic causal knowledge and specific information about the observed individual. Secondly, as I have already mentioned, although *individual*, claims about *statistical* individuals are nonetheless *generic*. Courgeau gives the reason himself: we search for a process underlying the *population*. Because they are representative of the population, individual random processes will give us *generic* causal knowledge.

6.3.4 Levels of Analysis and Variation in Multilevel Models

Among the different causal models presented in Chapter 3, multilevel analysis is by all means the most interesting for the present discussion. The reason simply is that it combines at once levels of analysis by aggregation and by discipline. For instance, Pickett and Pearl (2001: 111) see multilevel analyses of social factors and health as a possible reconciliation between two opposite paradigms: individual risk factor epidemiology and an ecological approach:

> Although there is an established tradition of inquiry into the impact of neighbourhood factors on sociological outcomes, such as educational attainment and labour market opportunities, neighbourhood variation in health has received less attention in epidemiology. In part this is because of the ecological fallacy when group level data are used to infer individual disease risks, and also because of a disciplinary focus on individual risk factors through much of this century. An increasing interest in societal influences on individual health status, along with improved statistical techniques for combining group level and individual level factors in regression models has spurred interest in contextual research in epidemiology.

Pickett and Pearl's critical review is of course relevant because it shows that multilevel analysis is a viable way out from the tension between holistic and individualistic approaches, i.e. in favour of levels of analysis by aggregation. But this is also relevant because multilevel analysis thus positively contributes to an analysis across fields: how social factors influence health outcomes. The relevance of modelling mixed mechanisms has already been discussed and I will not dwell on it longer.

I would like now to address the following question: is multilevel analysis prisoner of a new dualism, namely individual/social context? I will not answer this question directly; instead, I shall adopt a different strategy, namely discussing the rationale of variation in multilevel analysis. In particular, by comparing how the variation rationale works in classical aggregate or individual level models with multilevel models, we shall see whether it makes the individual/social context a prison or a fertile field for understanding society and human phenomena.

Let us consider, for instance, classical demographic analyses which assess the effect of a characteristic on a demographic behaviour. In one of the examples of Chapter 1, the goal was to establish the extent to which being a farmer affects migratory behaviour. Such analysis can be performed through standard regression techniques on aggregate characteristics. On the one hand, as Courgeau (2002) correctly points out, this is equivalent to Durkheim's methods of concomitant variations which has Millian origins, as we have seen in Chapter 4. The variation rationale is at this stage clearly at work and particularly this is a variation across characteristics (taxon 3 in section 4.2). Such an aggregate approach, however, is rather holistic. Durkheim, for instance, believed that there are collective tendencies (at the aggregate level) having existence independently of individuals. Causality is put at the societal level rather than in single-case relations, thus leading to the metaphysical danger of reifying aggregate variables into supervenient populational properties. Needless to say, there is also the methodological danger of falling into ecological fallacies such as we have already discussed. On the other hand, the specification of an individual model would lead to an equivalent regression function, except that parameters would be estimated from individual data and not from aggregate data. We are still within taxon 3 (see section 4.2), i.e. the interest is bestowed on variations of characteristics. We do not have particular metaphysical dangers here, although we have already came across the methodological ones, viz. atomistic fallacies.

I have already mentioned that multilevel analysis is an attempt to go beyond individualism and holism, and thus to avoid falling either in atomistic or ecological fallacies. Let us focus now on the idea of introducing contextual characteristics and of working simultaneously at the different levels. The goal is a more complex one: in explaining an individual behaviour (e.g. migratory behaviour), aggregate characteristics (e.g. percentage of farmers) are factors that influence this individual behaviour.

Thus, we have variation at the *individual level*—i.e. how the individual characteristics vary depending on another individual characteristic, for instance the tendency to migrate varies depending on being a farmer. We also have variation at the *contextual level*—i.e. how an individual characteristic varies depending on an aggregate characteristic, for instance the tendency to migrate varies depending on the

percentage of farmers present in a specific region. It is worth pointing out that aggregate models assume homogeneity of individuals *vis à vis* a given characteristic, for instance all farmers have the same probability to migrate. In multilevel analysis, the variation rationale allows us to question this assumption. Once we relax this homogeneity assumption, and we acknowledge instead a variation in the characteristic of interest (for instance the individual probability or tendency to migrate), the question becomes how this variation *varies* in different environments, i.e. at different levels of aggregation. In other words, the starting point will not be an assumed situation of homogeneity, but of *heterogeneity*, i.e. of *variation* across individuals with respect to a given characteristic.

To come back to the original question, we are now in a position to give a straight answer. No, multilevel analysis is not the prisoner of a new dualism individual/social context as long as the rationale of variation is adopted. In fact, the rationale of variation allows us not to be trapped in a rigid framework where the context determines individual behaviours. Instead, focusing on different sorts of variations (the 5 taxa of section 4.2), we can go back and forth from one level to another.

Chapter 7
Supporting the Rationale of Variation

Abstract This chapter supports the rationale of variation by showing how the notion of variation is involved or consistent with a number of philosophical accounts: the mechanist and counterfactual accounts, agency and manipulability theories, epistemic and singularist accounts.

Keywords Mechanist account; counterfactual account; agency theory; manipulability theory; epistemic causality; causality in single instances; W. Salmon; P. Dowe; D. Lewis; P. Menzies; H. Price; J. Woodward; D. Hausman; C. J. Ducasse; J. Williamson.

Introduction

Chapter 4 presented the rationale of variation at length and showed that this scheme of reasoning permeates causal models employed in the social sciences. As a matter of fact, in philosophy, a number of accounts of causality have been proposed in the last decades. The goal of this chapter is to disclose how this rationale is involved or, at least, is consistent with them.

In section 7.1 and section 7.2, I shall begin by discussing two traditional accounts, namely, the mechanist (Salmon-Dowe) and the counterfactual (Lewis) approach. Both approaches take causal relations to be 'objective', in the sense that causality is defined independently of the agent. In fact, the first relies on the notion of physical process, and the second on counterfactual logic. However, counterfactual reasoning also permeates causal modelling, although we shall see how far Pearl's account is from Lewis'. In section 7.3, I will look in the opposite direction, the one taken in the agency theory proposed by Menzies and Price. Causation is there defined relative to an agent's ability to operate on causal relations. The notion of manipulation thus becomes central for causality and has a highly anthropomorphic characterisation. In section 7.4, we shall see that the manipulability theory advanced by Hausman and Woodward develops a notion of intervention that fits causal modelling. This improvement tries to get rid of anthropomorphism and consequently

to regain objectivity lost in the agency approach. In section 7.5, I will examine a different attempt to depict causality *objectively*, namely Williamson's epistemic causality. The particularity of this approach is that 'objective' is there understood not as 'mind-independent', but as 'non-arbitrary'; however, objective causality, in the sense of physical causality, is denied. Section 7.6, instead, considers the question of whether the rationale of variation is also applicable in case a causal relation is only instantiated once. I shall attempt a positive answer by appealing to Ducasse's singularist account.

7.1 Variation in Mechanist Approaches

The mechanist approach has mainly been developed by Wesley Salmon (1984), subsequently criticised by Phil Dowe (1992), and then reformulated by Salmon in 1994.[1] Mechanist approaches take physical processes and interactions between them to be the fundamental concepts for causation. Those concepts are controversial, though. For convenience, I shall first sketch Salmon's original proposal, and then his restatement in reply to Dowe's criticism. We will see how the rationale of variation is involved in both versions.

In Salmon's theory, genuine causal processes are distinguished from *pseudo*-processes in that causal processes have the *ability* to transmit marks. In turn, the ability to transmit marks is spelled out by means of Reichenbach's *at-at theory* of mark transmission, which, briefly, requires the following (see Salmon 1984: 147–155; Reichenbach 1956: § 23): the transmission of a mark from point A in a causal process to point B in the same causal process is the fact that the mark appears at each point between A and B, without further interactions.

That is to say, when two processes intersect and display modifications that persist after the intersection, that interaction is called *causal* (Salmon 1984: 168–174). In other words, if the process is causal, a *change* or *variation* persists in the process after the causal interaction, according to the at-at theory. Notice, however, that being a causal process is only a necessary but not a sufficient condition for the production of lasting changes. Think for instance of two airplanes flying at different altitudes in different directions. These paths and their shadows on the ground may cross one another, but after the intersection both move on as if no such intersection had ever occurred. This happens because in this case processes are in fact *pseudo*-processes (Salmon 1984: 169).

Also, causal interactions involve a basic probability-raising requirement, typical of probabilistic theories of causality. Let C denote the intersection between two processes, let A denote the modification in one process and B the modification in the other, then the following inequality will generally hold:

$$P(A \cap B|C) > P(A|C) \cdot P(B|C). \tag{7.1}$$

[1] "Causality without counterfactuals" (1994) has been reprinted in Salmon (1998). In the following, I will refer to this second imprint.

This inequality conveys the idea that the *change* in both processes is *produced* by the interaction C, because, conditional on this interaction, the two modifications A and B turn out to be correlated.

This would be enough to show that the mechanist approach is definitively compatible with the variation rationale, if no objection had been raised against the requirement of positive statistical relevance. But such objections exist. Indeed, Salmon (1984: 200–202) himself discusses a possible counterexample. This is the so-called 'particle example' Suppose we have an atom in an excited state, say, at the 4th energy level. This atom may jump to the ground state in several ways, some of which involve intermediate occupation of the 1st level. Let $P\,(m \rightarrow n)$ stand for the probability that the atom in the mth level will make a direct transition to the nth level, and suppose that probabilities be allocated as follows:

1. $P\,(4 \rightarrow 3) = 3/4$.
2. $P\,(4 \rightarrow 2) = 1/4$.
3. $P\,(3 \rightarrow 1) = 3/4$.
4. $P\,(3 \rightarrow 2) = 0$.
5. $P\,(2 \rightarrow 1) = 1/4$.

It follows that the probability that the atom will occupy the 1st energy level in the process of jumping to the ground state is 10/16. However, if the atom occupies the 2nd level, then the probability of occupying the 1st level drops to 1/4, therefore occupation of the 2nd level is *negatively* relevant to occupation of the 1st level. Nevertheless, one would still consider the sequence 4th → 2nd → 1st level a causal chain. Salmon's reply is that, in spite of negative relevance, electron transition, in this fictitious example, would still be a *causal* process, since it satisfies requirements of at-at theory (Salmon 1984: 202).

It is worth noting that if the particle example troubles the mechanist approach, *a fortiori* it troubles the variation rationale too, for I just claimed compatibility between the two. I think that the objection to the variation rationale does not work, though. Let me explain why. In the first place, the objection does not really apply, for electron jumps belong to the domain of physics, and causality in physics falls beyond the scope of the present work. So, at most it might be argued against the applicability of the variation rationale in mechanist accounts, but not against the rationale *tout court*. Nonetheless, Salmon's reply leaves me room for a further defence. In fact, Salmon says that, although positive relevance fails, electron transition still satisfies the requirements of the at-at theory of causal influence, and since the at-at theory *plus* the notion of causal interaction have been showed to be compatible with the variation rationale, the particle example does not represent an overwhelming difficulty for my proposal.

However, as I mentioned, a more substantial critique of Salmon's account comes from Phil Dowe (1992). The strongest objection concerns the statistical characterisation of causal concepts: the notions of causal production and causal interaction cannot be analysed in terms of statistical relations. That is to say, the *statistical* formulation of conjunctive and interactive forks cannot adequately explicate *causal* concepts. Dowe thinks, instead, that an adjustment of Salmon's theory introducing

the concept of *conserved quantity* will do. The essence of Dowe's proposal is in two definitions:

(i) A causal interaction is an intersection of world-lines which involves exchange of a conserved quantity.
(ii) A causal process is a world-line of an object that manifests a conserved quantity.

A *conserved quantity*, that now plays the major role, is a quantity universally conserved, according to current scientific theories.

As a consequence of Dowe's critiques, Salmon then improved his theory. Three new definitions of (i) causal interaction, (ii) causal process, and (iii) causal transmission were worked out. Those definitions are supposed to make clearer Dowe's intuition that causal processes transmit conserved quantities. To plead for the variation rationale, in the new definition of causal interaction the notion of 'change' is again present and indeed plays a central role. So, the compatibility of mechanist accounts with the variation idea is not undermined; let me go over the argument in detail to explain why this is the case.

Originally, in Salmon's theory (Salmon 1984: 154–157), 'process' is a primitive notion not itself carrying any causal notion. Processes intersect with one another in space–time; however, space–time intersection, *per se*, is not a causal concept either. So Salmon needs a criterion to distinguish causal processes from pseudo-processes. Drawing on suggestions offered in Reichenbach (1956: § 23) about the mark method and causal relevance, Salmon comes to the conclusion that the mark method provides a good tool for accomplishing this task. Additionally, it was his view that what makes a process *causal* is its *capability* of transmitting marks, whether or not the process is actually transmitting one. This leads to a crucial counterfactual involvement in the characterisation of the mark method. Difficulties due to this counterfactual element have been raised by Cartwright, Kitcher, and, of course, Dowe. Salmon (1998: 253) then clarifies his point that the mark method ought to be regarded only as a useful experimental method for tracing or identifying causal processes, but it should not be used to *explicate* the very concept of causal process. Thus, in the restatement of his theory, Salmon abandons the mark criterion altogether and embraces Dowe's view that causal processes transmit conserved quantities, and *that* is what makes them causal. Let us now examine the three new definitions.

The first definition is borrowed from Dowe's theory. As Salmon (1998: 253) points out, Dowe's definition is equivalent to his principle of causal interaction, but much simpler:

Def$_1$: A causal interaction is an intersection of world-lines which involves exchange of a conserved quantity.

It is worth stressing the meaning of 'exchange': at least one outgoing process manifests a *change* in the value of the conserved quantity. The exchange is governed by the conservation law. Dowe (1992: 214) then points out that causes are related to conserved quantities and that those quantities—for instance, energy-mass, energy-momentum, and charge—are invariant. To which Salmon (1998: 254) replies that 'conserved' and 'invariant' quantities are not synonymous. A quantity is *conserved*

when its value does not change over time; in that case it is constant with respect to time translation. On the other hand, a quantity is *invariant* when it remains constant with respect to change of frame of reference. This distinction leads to subsequent reformulations of the definition of causal process; Def_2 below is the definitive one:

Def$_2$: A causal process is a world-line of an object that transmits a nonzero amount of an *invariant* quantity at each moment of its history (each space–time point of its trajectory).

Also, the following reformulation of the principle of mark transmission is proposed:

Def$_3$: A process transmits an invariant (or conserved) quantity from A to B ($A \neq B$) if it possesses this quantity at A and at B and at every stage of the process between A and B without any interactions in the half-open interval $(A, B]$ that involves an exchange of that particular invariant (or conserved) quantity.

Salmon concludes that although Dowe's conserved quantity theory embodies important improvements over the mark transmission theory, it is not fully satisfactory either. Instead, these three new definitions make considerable progress: they clean the definitions of causal interaction, causal transmission and causal process, on which to found a process theory of physical causality.

Indeed, it is a relevant improvement in the mechanist approach. Moreover, such progress reinforces my proposal, for the notion of variation still plays a fundamental role in the definition of causal interaction. In fact, an exchange of invariant (or conserved) quantities between processes actually produces a *modification* or *variation* in them, and this is what makes the interaction causal. Of course the way the rationale is here employed now differs from the quantitative one depicted in causal modelling, but the qualitative claim still holds: the bottom-line concept of causality is in the concept of variation, not in regularity, stability, or invariance.

7.2 Variation in Counterfactuals

David Lewis is the main proponent of the counterfactual theory of causation. In "Causation"[2] he develops an account where causal relations are analysed in terms of subjunctive conditionals, also called *counterfactuals*: 'A caused B' is interpreted as 'B would not have occurred if it were not for A'.

A counterfactual is a subjunctive conditional where the antecedent is known or supposed to be false. It can easily be shown that classical propositional logic does not fit the case of counterfactuals. In fact, if we were to analyse subjunctive conditionals as material implications, given that the antecedent is false, all counterfactuals would be equally true, according to the well known paradoxes of material implication. Therefore, Lewis explores a different path: counterfactuals are regimented by a *possible-worlds semantics*.

[2] "Causation", first published in 1973, has been reprinted in the collection *Philosophical papers* (1986, vol. II). In the following I will refer to this second imprint.

To begin with, possible-world semantics rests on the assumption of the existence of a plurality of worlds, among which there is also our actual world. This position is also known as *modal realism*. Secondly, worlds are compared with each other on the basis of their similarity or closeness. Worlds are thus ranged according to their similarity to our actual world: the closer the world is, the more similar it is to the actual world. To order worlds, we use a relation of *comparative over-all similarity*, which is taken as primitive; briefly, and very informally:

> A world w_1 is closer to our actual world w_a than another world w_2, if w_1 resembles to w_a more than w_2 does.

The truth of the counterfactual is then ascertained by an 'inspection' of what happens in other possible worlds. Given any two propositions A and B, the counterfactuals $A\Box\rightarrow B$ reads: 'if A were true, then B would also be true'. The counterfactual operator $\Box\rightarrow$ is defined by the following rule of truth:

The counterfactual $A\Box\rightarrow B$ is true (at a world w_i) if, and only if:

(i) There are no possible A-worlds;[3] or
(ii) Some A-world where B holds is closer to w than is any A-world where B does not hold.

The second case is the interesting one, for in the former the counterfactual is just vacuous. Notice, moreover, that in case A is true, the A-world is just our actual world and $A\Box\rightarrow B$ is true if, and only if, B is. But this is all about counterfactual logics. What about causation?

Lewis' theory is meant to apply to particular cases and not to causal generalisations. Also, causation is a relation between events—counterfactuals defined over propositions easily apply to events because, although presumably events are not propositions, they can at least be paired with them (Lewis 1986: 166). Causality comes in because by asking whether the counterfactual $A\Box\rightarrow B$ is true, we wonder whether B would be a consequence of the occurrence of A, i.e. whether the occurrence of A is the *cause* of the occurrence of B. So, the counterfactual, if true, states that if the cause had not occurred, then the effect would not have occurred either.

This is, very briefly, Lewis' theory. I shall not go through the technicalities of the counterfactual theory, nor shall I linger on criticisms. Lewis' hyper-realism, namely the assumption that all possible worlds are *real*, would be reason enough to abandon the theory altogether, or to search for more promising and sensible accounts of counterfactuals. In spite of this, I shall spend some time on the motivation that leads Lewis to develop his counterfactual theory. Once the motivation is spelled out, how the rationale of variation is involved in it will become apparent.

As is well known, the roots of the regularity analysis are in Hume's definition of causation in terms of constant conjunction and regular succession. Improvements with respect to Humean causation concern: (i) the distinction between causal laws and mere accidental regularities, (ii) causes and effects are subsumed under regularities by means of descriptions they satisfy, (iii) causes are roughly defined as any member of any minimal set of actual conditions that are jointly sufficient, given the

[3] In Lewis' theory 'A-world' means 'the world in which A is true'.

laws, for the existence of the effect. Those improvements are not enough for Lewis, though. Regularity accounts still face problems in distinguishing genuine causes from effects, epiphenomena, and pre-empted potential causes. Above all—Lewis argues—regularity accounts failed to notice the second definition of cause Hume gave in the *Enquiry concerning human understanding*:

> We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, *in other words*, where, if the first object had not been, the second never had existed. (*Enquiry*, Section VII, emphasis added.)

If the cause had not been, the effect had never existed. For Lewis, this is not just a restatement of the previous definition, but a clear input and encouragement to take the counterfactual path. In Lewis' words (Lewis 1986: 160–161):

> We think of a cause as something that makes a difference, and the difference it makes must a be a difference from what would have happened without it. Had it been absent, its effects—some of them, at least, and usually all—would have been absent as well.

Causes are supposed to make a difference, i.e. are responsible for *variations*. The bottom-line concept of causality is not in regularity, uniformity, or invariance, but is in difference, change, and variation. Agreed, it would be misleading to maintain that the counterfactual approach *relies* on the suggested rationale; yet, the variation idea is definitely consistent with it, as Lewis' words clearly show. The rationale of variation here involved is not quantitative at all, as is the case in causal modelling. Instead, there is a *qualitative* notion of variation at stake here. Surely Lewis' account is of little help in testing causality over a large data set, but counterfactuals do grasp, at least, our intuitions about how the causal relation works: *ceteris paribus*, if the cause had not occurred, the effect had never existed either. In other words, we expect the cause to be responsible for the *change* leading to the effect.

The case of Pearl's account of counterfactuals is different, however. In fact, Pearl (2000) attempts to evaluate counterfactuals by means of causal relations. So, counterfactuals do not explicate '*A* causes *B*' as in Lewis' account, and *a fortiori* do not need to be consistent with the variation rationale. Let me explain further.

Counterfactual reasoning is essential for science—e.g. for policy analysis or to predict the outcome of one experiment from the results of other experiments run under different conditions. Also, the formalization of causal inference calls for a language within which causal relationships in the world are distinguished from transitory relationships. For Pearl, structural models offer such distinction, and this is why a *structural model semantics*, rather than a possible-world semantics, is needed for counterfactuals. Pearl develops the structural model semantics and the subsequent axiomatization of counterfactuals in Chapter 7 of his (2000). Let us try to figure out how this semantics works.

As we saw in section 3.1.1, structural models consist of functional relationships among variables, and each relationship represents an autonomous mechanism. Counterfactuals are then defined in terms of the response of local modifications of those mechanisms. In other words, counterfactuals are "*derived* from (and in fact defined by) a functional causal model" (Pearl 2000: 35).

Consider again the canonical example of demand and price equilibrium we analysed in section 3.1.1. To recall, the model consists of two equations:

$$Q = \beta_1 P + \delta_1 I + \varepsilon_1 \tag{7.2}$$

$$P = \beta_2 Q + \delta_2 W + \varepsilon_2 \tag{7.3}$$

where Q is the quantity of households for a given product, P is the unit price, W is the wage rate, and ε_1, ε_2 represent errors. To put it very simply, a counterfactual question would ask: given that the current price is $P = p_0$, what would be the expected value of the demand Q if we were to control the price at $P = p_1$? To answer this question we have to consider the submodel:

$$Q = \beta_1 P + \delta_1 I + \varepsilon_1 \tag{7.4}$$

$$P = p_1 \tag{7.5}$$

The hypothetical value of price is replaced in the system of equations and this 'new' model is used to compute the expected value of the demand. It is worth emphasizing that in Pearl's account the evaluation of counterfactuals rests on the causal mechanisms involved in the structural model.

This brief presentation of Pearl's counterfactuals shows why, strictly speaking, counterfactuals do not need to be consistent with the rationale of variation, whereas their *structural* semantics does. But this work has already been done in section 4.1.

7.3 Variation in Agency Theories

The chief proponents of the agency theory are Peter Menzies and Huw Price (see Price 1991, 1992; Menzies and Price 1993). Agency theories analyse causal relations in terms of the ability of agents to achieve goals by manipulating their causes. In a nutshell, C is said to cause E if bringing about C would be an effective way to bring about E.

A number of objections have been raised against the agency approach. For instance, it has been argued that the notion of agency should not enter the analysis of causal claims, or that 'bringing about' is itself a causal notion, whence the indictment of circularity. Also, because causality thus defined is an anthropomorphic notion, it entails that causal relations would be different, were agents to be different, and at bottom, causal relations would not exist if agents did not exist either. In spite of this, Menzies and Price believe that the agency view is defensible and indeed carries more advantages than might be thought. In particular, these difficulties go away if we think of causation as a *secondary quality*. As dispositional theories of colour define it in terms of human capacities and responses, so does the agency theory for causality. In other words, the agency theory portrays causality as something analogous to a secondary quality, hence it should not be more controversial than the dispositional theory of colour.

This might vindicate the anthropomorphism bestowed on the theory, but a further point deserves discussion, namely what counts as an effective strategy for manipulating causes. Menzies and Price invoke the means–end relation, which is characterised in terms of *agent probabilities*. Agent probabilities are conditional probabilities assessed from an agent's perspective as follows: the agent probability of B conditional on A is the probability that should enter into the calculations of a rational agent, whose abilities consist in the capacity to realise or prevent A, and whose goals entirely concern B. So a strategy to bring about B is effective should a rational decision theory prescribe it as a way of bringing about B. Notice, however, that Menzies and Price disagree on the interpretation of 'agent probabilities'. Price seems to have a Bayesian conception, while Menzies holds that they are chances. That is to say, agent probabilities are defined in terms of their role in rational decision-making and this is why they embody a basis for a formal analysis of the means–end relation.

It is worth pointing out that their account is basically consistent with the central idea of probabilistic theories of causality. In fact, A constitutes a means for achieving B only in the case that the agent probability of $P_A(B)$ is greater than $P_{\text{not-}A}(B)$, where $P_A(B)$ denotes the agent probability that B would hold, were one to choose to realise A. Because the agency theory turns out to be consistent with probabilistic theories of causality, *a fortiori* it is consistent with the variation rationale. In fact, as we saw in section 4.1 statistical relevance relations aim at detecting *variations*.

In other words, the reason why the variation rationale is definitely compatible with the agency approach is that the effective strategy requires operating a *change* in the cause in order to produce or influence the effect. Also, because manipulative strategies are explicated in terms of rational decision-making procedures, the agency theory is consistent with the quantitative explication of the rationale by *measuring* variations.

I will not discuss further critiques of the agency theory, for which I direct the reader to Williamson (2005b). Williamson's criticisms mainly concern the following points. (i) The disagreement on the interpretation of agency probabilities makes unclear whether causality is conceived of as a physical relation or as a mental construct. (ii) Probabilistic dependence is shown to fail in many cases, for instance in quantum mechanics. Price's pragmatism (Price 2004) is also challenged. Instead, let me say a few words about the roots of the agency theory, because, if it is not already obvious that variation rationale is involved in the agency approach, it will become apparent shortly.

No doubt the intuition behind the Menzies–Price theory is not quantitative at all, but rather qualitative. Hurriedly in a footnote, Menzies and Price (1993: 187) say that their agency approach emerges from the forefathers of agency theories, namely Collingwood (1940), Gasking (1955) and von Wright (1971, 1975). I mention those authors because their tenets about causality also represent the background and the motivation for the Hausman-Woodward manipulability theory, which I will deal with in the next section. So, it will be worth spending some time on them.

At the beginning of his essay on causation, Collingwood says that the term 'cause' is ambiguous and has at least three senses: (i) historical, (ii) practical, and (iii) theoretical. The *first* one, the historical sense, refers to cases in which the cause

and the effect are human activities. This is worth noting because this is the sense that motivates an *agency* theory of causation (Collingwood 1940: 285):

> […] which is 'caused' is the free and deliberate act of a conscious and responsible agent, and 'causing' him to do it means affording him a motive for doing it.

The *second* significance says what sense of 'cause' is at stake when natural events are considered from a human point of view. Notice that this sense has a *practical* meaning, but not yet a theoretical one, which is instead reserved to the third sense of 'cause'. According to the practical sense, the primary aim is to enable man to enlarge his control of nature. Indeed, Collingwood says, this is the meaning of 'cause' that is used, for instance, in medicine or engineering. The third sense, the *theoretical* one, deals with events theoretically considered, i.e. as things that happen independently of each other and causation instead designates dependence between them. In truth, contemporary agency theorists conflate the first and second sense, and this is how agents, by manipulating causes, *bring about*, i.e. produce *variations*, in the effects.

Gasking (1955) at first excludes that causal priority and constant conjunction could successfully explicate causal relations and then suggests that, instead, *manipulation* will do: the notion of causation is thus highly connected to manipulative techniques for producing results (Gasking 1955: 483). His approach is also agency-dependent, for what Gasking has in mind, in the discussion of manipulative techniques, is how agents can produce events as means to producing subsequent events. Look now how the rationale of variation is involved. Here is the passage (Gasking 1955: 486):

> I have tried to give a general account of the producing-by-means-of relation itself: what is to produce *B* by producing *A*. We learn by experience that whenever we manipulate objects in a certain way a certain *change*, *A*, occurs. Performing this manipulation is then called: 'producing *A*'. We learn also that in certain special cases, or when certain additional conditions are also present, the manipulation in question also results in another sort of *change B*. In these cases the manipulation is also called 'producing *B*', and since it is in general the manipulation of producing *A*, in this case it is called 'producing *B* by producing *A*'. (My emphasis.)

The last source of motivation of agency and manipulationist theorists is Georg von Wright. In *Explanation and Understanding* (von Wright 1971) he devotes an entire section to causality and causal explanation, where he puts forward his understanding of causality within the philosophy of action. Again, we shall look for two ingredients, namely, agency-dependency and manipulation. A natural way of speaking of causes, says von Wright (1971: 64), is as factors which produce or bring about their effects and the way causes operate is compared to the operation of an agent who is responsible for what she does. Notably, causation is not tied to the mere notion of *doing* things, but rather with the notion of *bringing about* things and what we bring about are the effects of our action. Now the manipulative aspect comes in because "to say that we cause effects is not to say that agents are causes" (von Wright 1971: 69); instead, to say that we cause effects means that we do things that *themselves*, as causes, produce effects, namely *those* things act or operate as causes. Differently put (von Wright 1971: 70):

> [...] causality does not rest on an idea of doing things but itself provides a basis for possible manipulation.

And here comes von Wright's *experimentalist* notion of causation (von Wright 1971: 72):

> It is established that there is a causal connection between p and q when we have satisfied ourselves that, by manipulating the one factor, we can achieve or bring it about that the other is, or is not, there. We usually satisfied ourselves as to this by making experiments.

This emphasis on the fact that by manipulating causes we *bring about* effects, automatically renders the variation rationale—at least the *qualitative* claim—compatible with the agency version developed by von Wright.

So far so good. The agency theory as well as its forefathers' theories are definitively compatible and indeed rely on the notion of variation. Because Collingwood, Gasking and von Wright's accounts are also the groundwork for the manipulationist approach, the reader might think that she could skip directly to the section on epistemic causality, for the same arguments apply to the Hausman-Woodward approach and it would hence be a waste of time to go through the next paragraph. Let me instead recommend reading it, for further evidence in favour of the variation rationale is provided.

7.4 Variation in Manipulability Theories

Daniel Hausman and Jim Woodward propose a manipulationist account of causation. This approach is mainly developed in Hausman's *Causal asymmetries*, in Woodward's *Making things happen* and in two papers they published in the *British Journal for Philosophy of Science* (Hausman and Woodward 1999, 2004). Their purpose is to overcome the objection of anthropomorphism raised against the agency theory, by developing a notion of intervention which is not agency-dependent, and this reformulation is supposed to regain the objectivity of causal relations.

For the sake of philology, in spite of their agreement on the bulk of the manipulation view, Hausman and Woodward have rather different motivations. In Woodward's work, manipulation is an *explicandum* for causal explanation, whereas Hausman puts forward a manipulationist account specially to implement Menzies and Price's agency theory. I shall focus on the core of their agreement.

In their manipulationist or interventionist account, causal relationships have essentially two features: (i) they are potentially exploitable for purposes of manipulation and control and (ii) they are invariant under intervention. Everything turns around the specification of the notions of intervention and invariance. The notion of intervention is characterised in a nonanthropomorphic language, i.e. with no reference to human agency. The conditions that need to be met in an ideal experiment are the conditions under which a manipulation of the value of X is performed for the purpose of determining whether X causes Y. Briefly put, an intervention on X with respect to Y changes the value of X in such a way that, if any change occurs

in Y, it occurs only as a result of the change in the value of X and not from some other sources. Further assumptions are of course needed. Interventions should not be correlated with other causes of Y except for those causes of Y that are causally between the intervention and X or between X and Y; interventions should not affect Y *via* a route that fails to go through X. On the other hand, the notion of invariance is closely related to the notion of intervention and takes advantage of the notion of generalisation A generalisation G is invariant if it would continue to hold under some intervention that changes the value of X in such a way that the value of Y would change. 'Continue to hold' means that G correctly describes how the value of Y would change under this intervention. A generalisation G is the 'change-relating' relation we already came across in section 4.1.

Notice that the Hausman–Woodward account is not confined to cases in which the *actual* manipulation is possible, but is supposed to also fit cases where manipulation is impossible for practical or other sorts of reasons, for instance, in the explanation of past events, or in observational studies where manipulations are ethically unfeasible. This is why information relevant to manipulation needs to be understood modally or counterfactually. The modal or counterfactual understanding of causal relations raises other problems I shall not deal with; rather, I wish to stick to the notion of invariance.

For Hausman and Woodward, equations in structural equation models are correct descriptions of the causal relationship between X and Y if and only if were one to intervene in the right way to change the value of X, then Y should change in the way indicated by the equation. Now the key point is that the functional relation also has to be *invariant*, namely it should remain stable or unchanged under interventions. In other words, what invariance says is that the equation will continue to hold for some values of X, when those values are set by interventions (but only for a limited range of interventions).

Invariance, or structural stability, thus seems a necessary ingredient—although not the only one, as shown in section 4.3—to avoid the possibility that equations describe contingent or spurious relations, as for instance the correlation between sea levels in Venice and bread prices in England, or the correlation between birth rates and the number of storks in Alsace. Nonetheless, it is worth asking—and here is where the variation rationale emerges—*what* exactly remains invariant. So, invariance *of what*? The answer appears to be: invariance of a detected *variation*. As we saw in section 4.1, structural equations describe how variations in X accompany variations in Y and structural parameters quantify the causal relation between X and Y, and this is how the notion of variation is involved in them. Eventually, manipulability theories, in characterising causality as invariance under intervention, rely on the variation rationale in an essential manner. However, as discussed at length in section 4.1, it is conceptually wrong to locate the concept of causality in invariance.

Let me emphasise once more that invariance cannot be the bottom-line concept for causality. There are two reasons for this. In the first place, a further question about invariance can be asked, namely, *invariance of what*? Because variation turned out to be the answer, *variation* is the bottom-line concept. Secondly, as we saw in the discussion of the features of causal models (section 4.3), invariance is a specific

causal assumption and must be understood as a *constraint* on the variation, in order to allow the causal interpretation. To put it differently, to claim the primacy of variation is not trivial at all, because variation is a *pre*condition, in the sense that it *conceptually* precedes invariance, and *a fortiori* this entails the view that manipulability theories rely on the rationale of variation as well.

7.5 Variation in Epistemic Causality

So far, I have surveyed different accounts of causality; from the mechanist to the manipulationist approach, we witnessed a trade-off between objectivity and subjectivity. On the one hand, the mechanist approach patently takes causal relations to be objective, namely, *physically* 'out there', so to speak. On the other hand, the main objection raised against the agency theory is that their anthropomorphism makes causality highly subjective, whence the attempt of manipulability theories to work out a notion of intervention that is agency-independent and thus, again, objective. It seems that different senses of 'objective' and 'subjective' conflate, though. Williamson (2005a) starts afresh by developing an account where 'objective' means nonarbitrary; nonetheless, objective causality, in the sense of physical causality, is denied.

To begin with, in Williamson's approach, the border between the epistemology and the metaphysics of causality is clearly delineated. In the metaphysical account he advances, causality is a *mental* construct, and this mental construct is claimed to be objective. On the other hand, Williamson also gives an account of the epistemology of causality: causal relations are discovered by a hybrid of the inductive and hypothetico-deductive methods. Because the variation rationale belongs to epistemology, the question is whether or not the rationale is compatible with Williamson's learning strategies for epistemic causality. And indeed it is, as I will shortly show. Because Williamson's epistemology relies on his metaphysics, I shall deal with his metaphysics first.

Epistemic causality, as a mental construct, rests on the following two distinctions (Williamson 2005a: 8): subjective vs. objective, and mental vs. physical. *Mental* refers to features of an agent's mental state, whereas *physical* refers to features of the world 'out there'. Further, if two agents can differ as to causal structures, objectivity requires that at least one of them must be wrong. Williamson first develops his version of objective Bayesianism and claims that probabilities ought to be interpreted as mental *and* objective. That is, probabilities are interpreted as features of an agent's mental state. Objectivity requires that if two agents disagree as to probability values, then at least one of them must be wrong. See Williamson (2005a, b). Notice the slight shift in the terminology: traditionally, the literature calls 'objective' what Williamson calls 'physical', and 'subjective' what he calls 'mental'. It seems to me that Williamson's different terminology does better justice to a Bayesian interpretation of probability. To sum up, according to the epistemic view, causal relations are

mental rather than physical, although *objective* rather than subjective. Let me spell out this last point.

Causal relations belong to an agent's representation of the world, more precisely, epistemic causality deals with agents' *causal beliefs*. In fact, it is convenient to represent the world in terms of causes and effects, because such causal representations, if correct, enable predictions, diagnosis, decisions and interventions. Causality as a mental construct is thus made explicit. However, epistemic causality is not a matter of personal opinion, i.e. subjective. Epistemic causality is instead objective, in the sense that the causal interpretation ascribed to relations between variables is not *arbitrary*. Epistemic objectivity in fact requires that, *ceteris paribus*, if two agents disagree as to causal relationships, then at least one of them must be wrong. In other words, although causal relationships are mental constructs, they are not arbitrary.

Let us now turn to the epistemological side. Of course, the very possibility of determining causal relations *objectively* depends on the strategy for the discovery of causal relations. As I mentioned, Williamson proposes a hybrid between the inductive and the hypothetico-deductive methods. The core of this strategy does not substantially differ from the hypothetico-deductive method presented in section 3.2, but it is implemented with an inductive moment: the hypothesised causal structure that is to be confirmed or disconfirmed is *induced* from causal and probabilistic constraints imposed by background knowledge (Williamson 2005a: § 9.8–9.9).

The last step is to see whether or not the variation rationale is compatible with this strategy for learning causal relationships. And indeed it is. Let me explain why. Williamson's strategy for learning involves four stages:

1. Hypothesise.
2. Predict.
3. Test.
4. Update.

The *first* stage, the inductive one, requires a procedure for obtaining a causal graph from data; and the standard artificial intelligence techniques allow us to induce a minimal causal graph that satisfies the Causal Markov Condition, i.e. certain probabilistic independencies will hold in the induced causal graph. In the *second* stage, predictions are drawn from the induced graph and those predictions will be tested in the *third* stage. By renewed information or by performing experiments, predictions will be confirmed or disconfirmed. Of course, one would not expect predictions to be invariable consequences of causal claims; but, according to a probabilistic analysis, those predictions are supposed to hold in most cases. Finally, the *fourth* step represents a radical change in the hypothetico-deductive method: in case predictions fail, we do not start from the very beginning at step one, but we update the causal graph according to new evidence and information gathered.

Now the variation rationale permeates Williamson's strategy in the same sense that variation permeates causal models. In fact, a causal graph is a simplified version of a structural equation model, where only qualitative causal relationships are shown. As I discussed at length in section 4.1, according to the variation rationale,

causal models are tested by measuring suitable changes among variables. But this is exactly what happens in the test stage; witness Williamson (2005a: 152):

> If, for example, the model predicts that C causes E, and an experiment is performed which shows that intervening to change the value of C does not change the distribution of E, controlling for E's other direct causes, then this evidence alone may be enough to warrant removing the arrow from C to E in the causal model.

It seems to me that far from forcing and twisting others' thoughts around to my view, the variation idea marks a significant step in the clarification of fundamental concepts in the philosophy of causality.

7.6 Variation in Single Instances

Causal models try confirm or discover causal relations in large data sets. Consequently, the rationale of variation that emerges from causal modelling primarily concerns, at least in principle, repeatable observed phenomena. Two questions easily arise. Firstly, does it mean that causation in *single instances* is *ipso facto* ruled out? Secondly, if not, is the rationale of variation applicable in single instances?

As far as the first question is concerned, the epistemology of causality sketched in Chapter 4 does not exclude the possibility of causal relations that instantiate only once. In fact, the claim is not that *only* through a repetitive number of variations can causal relations be established; rather, in causal modelling, where, as a matter of fact, we deal with large correlational data, causal relations are tested by measuring suitable variations among variables of interest. In this sense, strictly speaking, single instances fall beyond the scope of the present work, but this does not entail that a sequence of events that occurs only once is not causal; differently put, single instances may well be causal.

Hence, we are led to discuss the second question: would the rationale of variation be applicable in single-case causal relations? C. J. Ducasse's account seems to provide a positive answer. In his 1926 article "On the Nature and Observability of the Causal Relation", Ducasse puts forward his *singularist* conception of causation. His approach rests on two main claims:

(i) A correct definition of 'causal relation' has to take into account the intuitive meaning of ordinary language, i.e. the definition of cause has to be formulated in terms of a *single* succession of certain events.
(ii) Thus defined, the causal relation is directly observable, *contra* Hume.

To see why the rationale of variation is compatible with this singularist account, let us consider Ducasse's causal *relata* and the definition of the causal relation. To begin with, Ducasse takes events, rather than Humean 'objects' to be the terms of the causal relation: it was the *hitting* of the vase with the hammer, and not the hammer itself, that caused its smashing. No doubt the characterisation and the status of events is a *vexata quaestio* of philosophy. However, what interests us here is the way Ducasse (1926: 58) conceives of events:

> […] an event is to be understood as a change or an absence of change (whether qualitative or relational) of an object.

More precisely, what changes, or remains unchanged, is a property of an object. For instance, it is the change of the position of the sun in relation to the building that caused the change of the shape of its shadow. So, as Psillos (2002: 67) puts it, it seems correct to say that Ducasse takes events to involve changes in the properties of things. Ducasse also allows for the distinction between type and token events. While type-events concern changes in objects generically understood, a token event specifies the particular object in which the change happens and also individuates the specific space–time location in which the change takes place. Let us now consider the definition of cause. It will be worth reading it fully (Ducasse 1926: 59):

> Considering two changes, *C* and *K* (which may be either of the same or of different objects), the change *C* is said to have been sufficient to, i.e. to have caused, the change *K*, if:
>
> 1. The change *C* occurred during a time and through a space terminating at the instant *I* at the surface *S*.
> 2. The change *K* occurred during a time and through a space beginning at the instant *I* and surface *S*.
> 3. No change other than *C* occurred during the time and through that space *C*, and no change other than *K* during the time through the space of *K*.
>
> More roughly, but in briefer and in more easily intuited terms, we may say that the cause of the particular change K was such particular change C as alone occurred in the immediate environment of K immediately before.

The notion of change—which is a qualitative synonym of the notion of variation presented in section 4.1—permeates Ducasse's view both of the characterisation of the *relata* as well as of the definition of the causal relation itself. What is more, in a paragraph below, Ducasse points out that the causal relation, thus defined, involves three terms, namely: (i) the environment of an object, (ii) some *change* in that environment, and (iii) the resulting *change* in the object.

It is worth noting that Ducasse is not denying the fact that there exist causal regularities in nature—such regularities exist and they are expressed by general statements linking event-types. Ducasse's point is that the hallmark of causation is *not* in regular succession, as the regularist view states. According to regularity theories, a particular sequence is said to be causal if it is an instantiation of a regularity, i.e. in case the token events *c* and *e* fall under the corresponding types *C* and *E* and in case all events of type *C* are regularly followed by events of type *E*. So, if the hallmark is not in regularity, it must be somewhere else—causation is to be fully definable only with reference to what happens in a single case.

Of course, several objections have been raised against Ducasse's proposal. For instance, it might be argued that such a definition is unable to offer a good mark for causation, because it offers only a sufficient condition for causation. Or, that the definition is vacuous, since it requires that for *C* to be the cause of *K*, *C* must be the only change in *K*'s environment before its occurrence; there is an epistemic

and an ontic difficulty there: on the one hand, how could we possibly identify this *single* change? On the other hand, what *exactly* was the single change in *K*'s environment that counts as the cause? I do not intend to challenge such criticisms, which are indeed well grounded. Nevertheless, it seems to me that Ducasse takes the right path in exploring the connections between the notions of *change* and of *causation*.

Concluding Remarks

1 Objectives, Methodology, and Results

Does quantitative social research take advantage of physical notions such as processes and interactions? Or does quantitative social research rather use statistical notions? But if so, do causal models aim at establishing regularities? And if this is the case, is it a rationale of regularity that governs causal modelling? If not, what rationale governs causal models in the social sciences? What notion guides causal reasoning in causal modelling?

This investigation into causal modelling has been twofold. On the one hand, by casting doubt on the dominant paradigm of regularity and proposing a new rationale of causality based on the notion of variation, my objective was to contribute to social sciences methodology. On the other hand, my objective was also to show that philosophy is not like Minerva's owl, and that *good* philosophical investigations into the scientific methodology and practice are indeed needed.

Causality as the measure of variation. In a nutshell, this is the rationale of causality I proposed. But by conceiving of causality *as* the measure of change, I did not provide a definition of what causality is. Instead, I gave a *rationale*. Epistemology and metaphysics ought not to be confused. Epistemological and metaphysical issues ought not to be conflated. Epistemological questions ought not to receive metaphysical answers.

A *rationale* is the principle or notion underlying some opinion, action, hypothesis, phenomenon, reasoning, model, or the like. The quest for a rationale of causality in causal modelling is the search for the concepts that guide causal reasoning and thanks to which we can draw causal conclusions. The search for the rationale of causality underlying causal models is the investigation of the concept of causality employed in those models, whether explicitly or implicitly. Consequently, to find a rationale of causality is, in the first place, a job for epistemology with important consequences for methodology. A *definition*, instead, is a description of a thing by its properties. A definition of causality states what causality *in fact is*. Some definitions attempt to define causality mind-independently; i.e. causality, whatever it is, is

F. Russo, *Causality and Casual Modelling in the Social Sciences: Measuring Variations* 211
Methodological Prospects in the Social Sciences

something physically 'out there'; causal relations, whatever they are, would be the same were agents different or even if there were no agents at all. Other definitions attempt to define causality mind-dependently, for instance by saying that causality is the agent's ultimate belief. In this case, constraints on the agreement of causal beliefs can be imposed in order to avoid arbitrariness, but causality is nonetheless in our minds and not 'out there', so to speak. Whether causality be 'physical' or 'mental', in either case a definition tries to say what causality *is* and, consequently, finding a definition of causality is a job for *metaphysics*. It is then clear that this investigation into causality and causal modelling is concerned with epistemology and methodology, viz. with the notions and the means that grant us epistemic access, that is the means that enable us to *know*, about causal relations.

The rationale I proposed is not the fruit of theoretical and abstract speculations, though. Instead, it emerges from a thorough analysis of the scientific practice and methodology. This is why I like to call the methodology of this investigation a *bottom-up* strategy, rather than top-down. We went bottom-up: from descriptive methodology to critical epistemology. From a very simple reading of case studies in Chapter 1, to a critical epistemology undertaken in the same chapter and carried on throughout this book. But we had a starting point. The descriptive task was in fact motivated by the dubious applicability of mechanist approaches to the social sciences. This descriptive task made our pivotal question—*What is the rationale of causality?*—more captivating, but it also left room for a number of seminal epistemological queries: *What guarantees the causal interpretation? How are levels of causation to be related? What is a causal mechanism? How is probability to be interpreted?*

We went bottom-up. Chapter 2 analysed probabilistic theories of causality as the philosophical attempt to provide a sensible scheme of the scientific practice. Although probabilistic theories failed to account for the multivariate aspect of causality in the social domain, they allowed us to isolate a number of key notions in a statistical analysis of causality, in particular, temporal priority of causes and statistical relevance. I also argued that to bring causal theory to maturity we need a thorough examination of causal models. This task was undertaken in Chapter 3. In that chapter, I presented the most widespread causal models used in the social sciences. These include structural equation models, covariance structure models, Granger-causality, Rubin's model, multilevel analysis and contingency tables. The ultimate objective of that chapter was to spell out the assumptions and the hypothetico-deductive methodology of causal models. Also, I offered a systematised exposition of the difficulties and weaknesses of causal modelling. The subsequent chapters were intended to provide a solution to the issues that had previously arisen, namely the rationale of causality and the guarantee of the causal interpretation (Chapter 4), the interpretation of probability in causal models (Chapter 5), the notion of mechanism and the relation between the levels of causation (Chapter 6).

We went up again. Chapter 4 presented an articulated epistemology of causality. First, I developed the rationale of causality based on the notion of variation rather than regularity or invariance. I have argued at length that causal models are tested by measuring variations, that is, *variation* is the bottom line concept of causality.

The rationale of causality is not given by the notions of regularity, nor of invariance. In particular, I argued that invariance and regularity are constraints to impose on the variation because variation conceptually precedes both of them, and that it is conceptually misleading to put causality into invariance or regularity. Second, I offered a taxonomy of variations causal models deal with. Third, I offered a vindication of the causal interpretation of the models analysed previously in Chapter 3. In fact, variation is not enough, by itself, to guarantee a causal interpretation. This is why we carried out a thorough analysis of causal models, and it turned out that the causal interpretation is granted thanks to an extremely rich apparatus that consists of assumptions and background knowledge.

Having reached the climax of the philosophical investigation, we then went back to methodology, and particularly to the methodological consequences of the epistemology offered. Chapter 5 discussed at length the interpretation of probability and argued in favour of a Bayesian interpretation. The argument was structured in two steps. I firstly defended both an empirically-based and an objective Bayesian interpretation on the grounds that they allow a sensible interpretation of both generic and single-case probabilistic causal claims. Secondly, I showed the net gain of adopting the objective Bayesian interpretation in causal modelling. I argued that scientists would be better off with an objective Bayesian interpretation as it enlightens the design and interpretation of tests, and as it is a better guide for social policy and decision making.

Chapter 6 addressed the last methodological issues. Firstly, I attempted a characterisation of causal mechanisms and argued that (i) causal modelling ought to be the modelling of causal mechanisms, (ii) modelling mechanisms participates in the explanation of a given phenomenon. Secondly, I cast doubts on metaphysical accounts that assume different mechanisms operating at different levels of causation. We then saw how levels of causation are to be *epistemologically* related. The reformulation of the Connecting Principle made clear the idea that causal conclusions drawn from causal models primarily concern the population and that these generalisations are used to assess the *likelihood* of single-case hypotheses, but not the strength of individual-level causal relations. Thirdly, I tried to straighten out the debate on the levels of causation. I showed that philosophers and scientists tackle rather different questions with a fairly similar terminology, whence the source of confusion. I also argued that the problem of the levels of causation has to be reformulated as a problem of levels of analysis and I assessed the extent to which multilevel models provide a good framework.

Finally, Chapter 7 was meant to offer further theoretical support to the rationale of variation by showing that the concept is latent in many accounts of causality, but never made explicit. We saw the primary role played by the notion of variation in mechanist and counterfactual accounts, in agency and manipulability theories, in the epistemic approach and in single instances.

Let me emphasise once again: this is all about methodology and epistemology. But what about the metaphysics of causation? On purpose, metaphysical issues were left aside. I mentioned that some of the assumptions of causal models are of course metaphysical in character—for instance causal priority—although their plausibility

or correctness has not been challenged. In the discussion of the levels of causation I raised doubts about the plausibility of postulating different mechanisms on the ground of parsimony arguments. But this remark had only methodological and epistemological significance, rather than metaphysical significance. In the ultimate analysis, no stance has been taken concerning causal realism, although I said a few words on this issue in the Introduction.

2 The Methodological Import of Philosophical Results

We finally come to the very last section. Throughout the book, I analysed quantitative social science methodology. Yet, the first output is *philosophical*. I cast doubt on the received view which is based on a rationale of regularity and I advanced, instead, the view that causal modelling is governed by a rationale of variation, not of regularity nor of invariance. The notion of variation has been unduly neglected both in the philosophical and in the scientific literature. The burden of regularity, a heritage of Hume, influenced philosophers as well as scientists, leading (i) to conceptual mistakes—e.g. causality is in invariance—and (ii) to cumbersome concepts—e.g. exceptionless laws that do not fit social contexts—and (iii) to over-sophisticated methods hiding the right rationale of causality. But as I said, the variation rationale is not an intellectual intuition to be imposed on practising scientists, nor is it the result of a mere literature review. It arises from a bottom-up methodology of research in philosophy, namely a philosophical investigation that *starts* from the scientific practice, *within* the scientific practice raises methodological and epistemological issues, and *for* the scientific practice points to the path forward. The second result is then methodological. In the remainder of this section I recall the methodological gains of taking seriously the epistemology of causal modelling offered in Chapter 4.

Correlation does not prove causation. Regularity does not prove causation either. Causal models, as discussed in section 4.3, have a richer statistical apparatus than simple associational models—this is what justifies the causal interpretation. Structural stability, no confounding, covariate sufficiency are the key causal assumptions. But can we improve the quality of these assumptions in order to better guarantee the causal interpretation?

According to the hypothetico-deductive methodology of causal models, causal hypotheses are first formulated and then put forward for empirical testing. In turn, empirical testing will tell us whether the hypothesised causal structure has to be accepted or rejected. The discussion of the H-D methodology allowed us to emphasise the central role played by background knowledge in the process of model building and of testing. However, as I will try to show next, to improve the quality of causal assumptions we must adopt the rationale of variation.

Why should we adopt the variation rationale then? In fact, as my point is that the variation rationale is *already* employed (sometimes not explicitly, though) in causal modelling, the question rather turns into: why should the practising scientist acknowledge the use of this rationale? The answer is that the practising scientist will

then look for the right thing, that is variation, not regularity. The practising scientist will then build and test the causal model without the bias of regularity. Regularity refers to the possibility of finding universal sequels of events, and, perhaps, universal laws in social domains too. But this has been exactly the main reason why it is often claimed that the social sciences cannot reach the kingdom of the hard sciences. The rationale of variation thus relieves the social sciences from cumbersome laws and offers a more flexible and appropriate tool to describe causal relations in social contexts.

The rationale of variation also gives us a better understanding of invariance vs. regularity. The invariance condition does not ensure the presence of a regular succession of events *à la* Hume, but of a *joint variation* which is stable under a class of environmental changes. Stability is nonetheless local. It is confined to the social context under analysis, and therefore to the causal model. However, this is a virtue rather than a drawback. The objects of the social sciences vary too much across time and space, therefore (i) a general stability is virtually impossible to find, and (ii) even it were possible, those stable relations would be very approximate and inaccurate, and thus of scarce utility for understanding social phenomena, let alone for policy reasons. But the rationale of variation can do even more than that, as it enlightens no-confounding and covariate sufficiency too. Together, those two assumptions imply that all and only the variables in the model play a causal role. What is exactly this role? Has this role something to do with regularity? It does not. Following Wunsch (2007), a confounding variable is a common cause both of the putative cause and of its outcome. To be more precise, in Reichenbachian terms, the confounder screens off the cause from the effect, or it screens off the *variation* in the effect due to the *variation* in the cause. Screening off thus concerns *variations*, although it also has fatal consequences for structural stability (Mouchart et al. 2008).

A further methodological gain in explicitly adopting the variation rationale is related to the action-oriented goal of social research. The rationale of regularity offers little, if any help at all, when it comes to set social policies. The rationale of variation, instead, proves to be useful. The goal of social policies is to intervene on society in order to change a given state of affairs. Interventions have to be planned depending on what *variations* hold, because it is on the relata of the variation that we have to intervene, not on the regular succession of events. Although causal variations typically require structural stability, they do not necessarily need Humean regularity. For instance, in epidemiological contexts a plausible physiological mechanism can provide good evidence. On this point, see Kundi (2006). For a discussion on the types of evidence required in the health sciences, see Russo and Williamson (2007c).

However, in the social sciences we do not have physiological mechanisms, but the socio-demo-economic mechanism is exactly what we want to model by means of variations, as it has been discussed at length in Chapter 6.

As I discussed in Chapter 6, different variables store different sorts of information. The causal story has to be coherent on ontological grounds and on epistemic grounds—the causal mechanism thereby modelled has to be useful for action-oriented goals beside the cognitive goals. Ontological coherence is achieved by selecting the relata of the causal variation at the same level of reality or by specifying a

suitable mechanism. For instance, social variables do not cause variations in health variables, unless we specify the mechanism responsible for that. Epistemic coherence is achieved by taking into account background knowledge. We need results to be consistent with available scientific theories. The socio-demo-political-economic context not only provides the causal framework—and thus participates into the causal interpretation—but it also constitutes the sounding board of causal relations. Finally, the causal mechanism is useful for the cognitive goal if it provides a causal explanation of the phenomenon under study and if it informs policies about what variables one has to intervene upon.

That causal relations have to be probabilistically modelled is not very controversial. Disagreement comes as to how those probabilities are to be interpreted. This might seem an innocuous issue, but in fact it is not. The choice of the interpretation of probability has important consequences for epistemology, methodology and metaphysics. For instance, in Russo (2006a) I discuss the implications of the interpretation of probability for theoretical realism in the physical sciences. Russo and Williamson (2007b) tackle the problem of the interpretation of probability in cancer science and argue that a twinned frequentist-cum-objective Bayesian interpretation is needed to account for different types of causal claims. Here, I shall confine the discussion to some general methodological and epistemological considerations. In Chapter 5, I made the case for a Bayesian interpretation of probability, in particular for an objective Bayesian interpretation. This interpretation is preferable on several grounds. On the one hand, Bayesian interpretations offer a better understanding of test statistics and allow us to evaluate the probability of a hypothesis, which is not possible in the framework of classical statistics. On the other hand, an objective Bayesian interpretation accommodates causal claims made at different levels of causation (the generic vs. single case distinction) and offers a better ground for policy making. Despite the fact that Bayesian accounts interpret probability as rational degrees of belief, objective Bayesian probabilities are not devoid of empirical content and can also explain the widespread use of frequencies in science, as frequencies provide evidence upon which we shape degrees of belief.

Anti-causal prophecies have been disproved. Causality is neither a "relic of a bygone age", nor "another fetish of modern science". It has never lost its appeal. On the contrary, in its long history as a central topic of philosophy, causality continues to show fascinating interconnections with germane issues, as for instance explanation, interpretation of probability, scientific realism, ontology, and so forth. It is at once the complexity and the beauty of causality—that it forces us to take a stance on these several other problems. I would mainly stress its beauty, though. Because it is in their multi-disciplinary aspect that causal questions lead research onward to a better understanding of nature and society, as we lay a vast causal mosaic, filled tile after tile.

Bibliography

Abbott, A. (1998). The causal devolution. *Sociological Methods and Research*, 27(2), 148–181.

Adams, P., Hurd, M., McFadden, D., Merrill, A., Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, 112, 3–56. With discussion.

Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.

Aish-Van Vaerenbergh, A. -M. (1994). Modèles statistiques et inférences causales: analyse des structures de covariances avec LISREL. In R. Franck (Ed.), *Faut-il chercher aux causes une raison?* (pp. 106–130). Paris: Vrin.

Aish-Van Vaerenbergh, A. -M. (2002). Explanatory models in suicide research: explaining relationships. In R. Franck (Ed.), *The explanatory power of models* (pp. 51–66). Dordrecht: Kluwer.

Aristotle (1928). *Metaphysics*. Translated by W. D. Ross, *The works of Aristotle translated into English*, vol. 8, Second edition. Oxford: Clarendon.

Atmanspacher, H., Bishop, R. (2002) (Ed.). *Between chance and choice: interdisciplinary perspectives on determinism*. Thoverton: Imprint Academic.

Bayarri, S., Berger, J. O. (2000). P-values for composite null models. *Journal of the American Statistical Association*, 95(452), 1127–1142. With discussion.

Beck, W., van der Maeson, L., Walker, A. (2001). Introduction: who and what is the European Union for? In W. Beck, L. van der Maeson, G. Thomése, A. Walker (Ed.), *Social quality: a vision for Europe* (pp. 1–18). The Hague: Kluwer Law International.

Berger, J. O., Boukai, B., Wang, Y. (1997). Unified frequentist and Bayesian testing of a precise hypothesis. *Statistical Science*, 12(3), 133–160.

Berger, J. O., Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 3(2), 317–335.

Berger, J. O., Guglielmi, A. (2001). Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives. *Journal of the American Statistical Association*, 96(453), 174–184.

Berkovitz, J. (2002). On causal inference in determinism and indeterminism. In H. Atmanspacher, R. Bishop (Ed.), *Between chance and choice: interdisciplinary perspectives on determinism* (pp. 237–278). Thoverton: Imprint Academic.

Bernardo, J. M. (1980). A Bayesian analysis of classical hypothesis testing. In J. Bernardo, M. H. DeGroot, D. V. Lindley, A. F. M. Smith (Ed.), *Bayesian statistics* (pp. 605–647). Valencia: Valencia University Press. With discussion.

Bessie, J. D. (1993). On the strength of a causal chain. *Pacific Philosophical Quarterly*, 74(1), 11–36.

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Blalock, H. M. (1968a). The measurement problem: a gap between the languages of theory and research. In H. M. Jr. Blalock, A. Blalock (Ed.), *Methodology in social research* (pp. 5–27). New York: McGraw Hill.

Blalock, H. M. (1968b). Theory building and causal inferences. In H. M. Jr. Blalock, A. Blalock (Ed.), *Methodology in social research* (pp. 155–198). New York: McGraw Hill.

Boudon, R. (1967). *L'analyse mathématique des faits sociaux*. Paris: Plon.

Boudon, R., Lazarfeld, P. (1966) (Ed.). *L'analyse empirique de la causalité*. Paris: Mouton & Co.

Bovens, L., Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Clarendon Press.

Bunge, M. A. (1979a). *Causality and modern science*. New York: Dover. Third revised edtion.

Bunge, M. A. (1979b). *A world of systems*. Dordrecht: Reidel Publishing Company.

Bunge, M. A. (1996). *Finding philosophy in social science*. New Haven, CT: Yale University Press.

Bunge, M. A. (1998). *Social science under debate: a philosophical perspective*. Toronto: University of Toronto Press.

Bunge, M. A. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences*, 34(2), 182–210.

Caldwell, J. C. (1979). Education as a factor in mortality decline: an examination of Nigerian data. *Population Studies*, 33(3), 395–413.

Campbell, D. T., Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Carnap, R. (1951). *The logical foundations of probability*. Chicago: University of Chicago Press. Second edition.

Carroll, J. W. (2003). Laws of nature. In E. N. Zalta (Ed.), *The Stanford encyclopaedia of philosophy*. (Fall 2003 edition). http://plato.stanford.edu/archives/fall2003/entries/laws-of-nature/. Accessed 12 March 2008.

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs*, 13(4), 419–437.

Cartwright, N. (1989). *Nature's capacities and their measurements*. Oxford: Clarendon Press.

Cartwright, N. (1995). Précis of 'Nature's capacities and their measurement'. *Philosophy and Phenomenological Research*, 55(1), 153–156.

Cartwright, N. (1997). What is a causal structure? In V. McKim, S. P. Turner (Ed.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 343–357). Indiana: University of Notre Dame Press.

Cartwright, N. (1999). Causal diversity and Markov condition. *Synthese*, 121, 3–27.

Cartwright, N. (2000). Measuring causes: invariance, modularity and the causal Markov condition. *Measurement in physics and economics discussion paper series*. Centre for Philosophy of Natural and Social Science, 10/00.

Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two: comments on Hausman and Woodward. *British Journal for the Philosophy of Science*, 53(3), 411–453.

Cartwright, N. (2007a). *Hunting causes and using them: approaches in philosophy and economics*. Cambridge: Cambridge University Press.

Cartwright, N. (2007b). Are RCTs the gold standard? *Bio Societies*, 2, 11–20.

Chakravartty, A. (2005). Causal realism: events and processes. *Erkenntnis*, 63, 7–31.

Cohen, J. L. (1989). *An introduction to the philosophy of induction and probability*. Oxford: Clarendon Press.

Collingwood, R. (1940/1998). *An essay on metaphysics*. Oxford: Clarendon Press.

Cook, T. D., Campbell D. T. (1979). *Quasi-experimentation. Design and analysis issues for field settings*. Chicago: Rand MacNally.

Cook, T. D., Campbell D. T. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 68, 141–180.

Corfield, D., Williamson J. (2001) (Ed.). *Foundations of Bayesianism*. Dordrecht: Kluwer.

Courgeau, D. (1994). Du groupe à l'individu: l'exemple des comportements migratoires. *Population*, 1, 7–26.

Courgeau, D. (2002). Vers une analyse biographique multiniveaux. In M. Christine (Ed.), *Actes des journées de méthodologie statistique* (pp. 375–394), INSEE méthodes 101.

Courgeau, D. (2003) (Ed.). *Methodology and epistemology of multilevel analysis. Approaches from different social sciences*. Dordrecht: Kluwer.

Courgeau, D. (2004a). *Du groupe à l'individu: synthèse multiniveau*. Paris: Editions de l'INED.

Courgeau, D. (2004b). Probabilité, démographie et sciences sociales. *Mathématiques et sciences humaines/Mathematics and Social Sciences*, 167(3), 27–50.

Courgeau, D. (2007a). *Multilevel synthesis: from the group to the individual*. Dordrecht: Springer.

Courgeau, D. (2007b). Inférence statistique, échangeabilité et approche multiniveau. *Mathématique et sciences humaines/Mathematics and Social Sciences*, 179(3), 5–19.

Cox, D. R. (1992). Causality: some statistical aspects. *Journal of the Royal Statistical Society*, 155(2), 291–301.

Cox, R. T. (1946). Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14(1), 1–13.

Cox, D. T. (2000). Causal inference without counterfactuals: commentary. *Journal of the American Statistical Association*, 95, 424–425.

Dawid, A. P. (1982). The well-calibrated Bayesian. *Journal of the American Statistical Association*, 77(379), 605–610.

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95, 407–427.

Dawid, A. P. (2002a). Influence diagrams for causal modelling and inference. *International Statistical Review*, 70, 161–189. Corrigenda, ibidem, 437.

Dawid, A. P. (2002b). Commentary: counterfactuals: help or hindrance? *International Journal of Epidemiology*, 31, 429–430.

Dawid, A. P. (2007). Counterfactuals, hypotheticals and potential responses: a philosophical examination of statistical causality. In F. Russo, J. Williamson (Ed.), *Causality and probability in the sciences* (pp. 503–532). London: College Publications.

de Finetti, B. (1937). Foresight. Its logical laws, its subjective sources. In H. E. Kyburg, H. E. Smokler (Ed.), *Studies in subjective probability* (pp. 194–207). Huntington, New York: Wiley.

de Finetti, B. (1993). *Probabilità e induzione. Induction and probability*. Edited by D. Montanari, D. Cocchi. Bologna: CLUEB.

Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science*, 59, 195–216.

Drèze, J. H., Mouchart, M. (1990). Tales of testing Bayesians. In R. A. L. Carter, J. Dutta, A. Ullah (Ed.), *Contributions to econometric theory and applications - Essays in honour of A. L. Nagar* (pp. 345–366). New York: Springer.

Droesbeke, J.-J., Fine, J., Saporta, G. (2002). *Méthodes bayésiennes en statistique*. Paris: Technip.

Ducasse, J. C. (1926). On the nature and observability of the causal relation. *Journal of Philosophy*, 23, 57–68.

Duchêne, J., Wunsch, G. (1985). From theory to statistical model. In *IUSSP, International Population Conference* (pp. 209–224), Volume 2. Liège: Ordina.

Duchêne, J., Wunsch, G. (1989). Conceptual frameworks and causal modelling. In L. Ruzicka, G. Wunsch, P. Kane (Ed.), *Differential mortality. Methodological issues and biosocial factors* (pp. 21–35). Oxford: Clarendon.

Duchêne, J., Wunsch, G. (1989) (Ed.). *L'explication en sciences sociales: la recherche des causes en démographie*. Chaire Quetelet 1987. Louvain-la-Neuve: CIACO.

Duchêne, J., Wunsch, G. (2006). Causalité et modèles causaux. In G. Caselli, J. Vallin, G. Wunsch (Ed.), *Démographie: analyse et synthèse* (pp. 315–334), Volume VIII, Observation, méthodes auxiliaires, enseignement et recherche. Paris: Editions de l'Institut National d'Etudes Démographiques.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Dupré, J., Cartwright, N. (1988). Probability and causality: why Hume and indeterminism don't mix. *Noûs*, 22(4), 521–536.
Durkheim, E. (1895/1912). *Les règles de la méthode sociologique*. Paris: Libraire Félix Arcan. Sixth edition.
Durkheim, E. (1897/1960). *Le suicide*. Paris: Presses Universitaires de France.
Edwards, A. W. F. (1972). *Likelihood. An account of the statistical concept of likelihood and its application to scientific inference*. Cambridge: Cambridge University Press.
Edwards, A. W. F. (1997). What did Fisher mean by 'inverse probability' in 1912–1922? *Statistical Science*, 12(3), 177–184.
Eells, E. (1991). *Probabilistic causality*. Cambridge: Cambridge University Press.
Eells, E., Sober, E. (1983). Probabilistic causality and the question of transitivity. *Philosophy of Science*, 50, 35–57.
Ehring, D. (1984). Probabilistic causality and pre-emption. *British Journal for the Philosophy of Science*, 35, 55–57.
Ellett, F. S., Ericson, D. P. (1983). The logic of causal methods in social science. *Synthese*, 57, 67–82.
Ellett, F. S., Ericson, D. P. (1984). Probabilistic causal systems and the conditional probability approach to causal analysis. *Quality and Quantity*, 18, 247–259.
Ellett, F. S., Ericson, D. P. (1986a). Correlation, partial correlation, and causation. *Synthese*, 67, 157–173.
Ellett, F. S., Ericson, D. P. (1986b). An analysis of probabilistic causation in dichotomous structures. *Synthese*, 67, 175–193.
Ellett, F. S., Ericson, D. P. (1989). Causal modelling and theories of causation. In J. Duchêne, G. Wunsch (Ed.), *L'explication en sciences sociales: la recherche des causes en démographie* (pp. 397–424). Chaire Quetelet 1987. Louvain-la-Neuve: CIACO.
Engle, R. F., Hendry, D. F., Richard, J. -F. (1983). Exogeneity. *Econometrica*, 51(2), 277–304.
Fetzer, J. (1988) (Ed.), *Probability and causality*. Dordrecht: Reidel.
Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, 222, 309–368.
Fisher, R. A. (1925). *Statistical methods for research workers*. London: Oliver and Boyd. http://psychclassics.yorku.ca/Fisher/Methods/. Accessed 14 March 2008.
Fisher, R. A. (1930). Inverse probability. *Proceeding of the Cambridge Philosophical Society*, 26, 528–535.
Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society*, 98(1), 39–82.
Florens, J. -P., Mouchart, M. (1989). Bayesian specification tests. In B. Cornet, H. Tulkens (Ed.), *Contributions in operations research and economics* (pp. 467–490). Cambridge: MIT Press.
Florens, J. -P., Mouchart, M. (1993). Bayesian testing and testing Bayesians. In G. S. Maddala, C. R. Rao (Ed.), *Hand-book of statistics* (pp. 303–334). Amsterdam: North Holland.
Franck, R. (1994) (Ed.). *Faut-il chercher aux causes une raison?* Paris: Vrin.
Franck, R. (1995). Mosaïques, machines, organismes et sociétés. *Revue Philosophique de Louvain*, 93(1–2), 67–81.
Franck, R. (2002) (Ed.). *The explanatory power of models*. Dordrecht: Kluwer.
Franck, R. (2003). Causal analysis, systems analysis, and multilevel analysis: philosophy and epistemology. In D. Courgeau (Ed.), *Methodology and epistemology of multilevel analysis. Approaches from different social sciences* (pp. 175–198). Dordrecht: Kluwer.
Freedman, D. A. (1999). From association to causation: some remarks on the history of statistics. *Statistical Science*, 14(3), 243–258.
Freedman, D. A. (2004a). Statistical models for causation. Technical Report 651. http://www.stat.berkeley.edu/~census/651.pdf. Accessed 14 March 2008.
Freedman, D. A. (2004b). On specifying graphical models for causation, and the identification problem. *Evaluation Review*, 26, 267–93.
Freedman, D. A. (2005). *Statistical models. Theory and practice*. Cambridge: Cambridge University Press.

Freedman, D., Pisani, R., Purves, R. (1998). *Statistics*. New York: W.W. Norton. First edition.
Fried, H. O, Schmidt, S. S., Lovell, K. (1993) (Ed.). *The measurement of productive efficiency*. New York: Oxford University Press.
Galavotti, M. C., Suppes, P., Costantini, D. (2001) (Ed.), *Stochastic causality*. Stanford: CSLI.
Gasking, D. (1955). Causation and recipes. *Mind*, 64, 479–487.
Gérard, H. (1989). Théories et théorisation. In J. Duchêne, G. Wunsch (Ed.), *L'explication en sciences sociales. La recherche des causes en démographie* (pp. 267–281). Chaire Quetelet 1987. Louvain-la-Neuve: CIACO.
Gérard, H. (2006). De la théorisation en démographie. In G. Caselli, J. Vallin, G. Wunsch (Ed.), *Démographie: analyse et synthèse* (pp. 291–314). Volume VIII, Observation, méthodes auxiliaires, enseignement et recherche, Paris: Editions de l'Institut National d'Etudes Démographiques.
Giere, R. (1984). Causal models with frequency dependence. *Journal of Philosophy*, 81, 384–391.
Giere, R. (1999), *Science without laws*. Chicago: University of Chicago Press.
Gillies, D. (2000). *Philosophical theories of probability*. London: Routledge.
Glymour, C., Scheines, R. (1986). Causal modeling with the TETRAD program. *Synthese*, 68, 37–64.
Glymour, C., Scheines, R., Spirtes, P., Kelly K. (1987). *Discovering causal structure: artificial intelligence, philosophy of science, and statistical modelling*. San Diego: Academic.
Goldberger, A. S. (1972). Structural equation methods in the social sciences. *Econometrica*, 40, 979–1001.
Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.
Goldstein, H. (2003). *Multilevel statistical models*. Kendall's Library of Statistics, 3. London: Arnold.
Good, I. J. (1959). A theory of causality. *British Journal for the Philosophy of Science*, 9, 307–310.
Good, I. J. (1961). A causal calculus I. *British Journal for the Philosophy of Science*, 11, 305–18. Reprinted in I. J. Good, *Good thinking. The foundations of probability and its applications* (pp. 197–217). Minneapolis: University of Minnesota Press.
Good, I. J. (1962), "A causal calculus II", *British Journal for the Philosophy of Science*, 12, pp. 43–51. Reprinted in I. J. Good, *Good thinking. The foundations of probability and its applications* (pp. 197–217). Minneapolis: University of Minnesota Press.
Good, I. J. (1972). Review of Patrick Suppes 'A Probabilistic Theory of Causality'. *Journal of American Statistical Association*, 67(337), 245–246.
Good, I. J. (1977). Explicativity: a mathematical theory of explanation with statistical applications. *Proceedings of Royal Society*, A 354, 303–330. Reprinted in I. J. Good, *Good thinking. The foundations of probability and its applications* (pp. 219–236). Minneapolis: University of Minnesota Press.
Good, I. J. (1980). Some comments on probabilistic causality. *Pacific Philosophical Quarterly*, 61, 301–304.
Good, I. J. (1983a). *Good thinking. The foundations of probability and its applications*. Minneapolis: University of Minnesota Press.
Good, I. J. (1983b). The philosophy of exploratory data analysis. *Philosophy of Science*, 50, 283–295.
Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438.
Grawitz, M. (1996). *Méthodes des sciences sociales*. Paris: Dalloz. Tenth edition.
Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, 29, 722–729.
Grünbaum, A. (1963). *Philosophical problems of space and time*. New York: Alfred A. Knopf.
Gutiérrez-Fisac, J. L., Regidor, E., Banegas Banegas, J. R., Rodriguez Artalejo, F. (2002). The size of obesity differences associated with educational level in Spain, 1987, and 1995/97. *Journal of Epidemiology and Community Health*, 56, 457–460.
Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, 12, iii–vi + 1–115.

Hacking, I. (1965). *Logic of statistical inference*. London: Cambridge University Press.
Hacking, I. (1978). *The emergence of probability: a philosophical study of early ideas about probability, induction and statistical inference*. Cambridge: Cambridge University Press.
Hage, J., Meeker, B. F. (1988). *Social causality*. Boston: Unwin Hyman.
Hájek, A. (2003). Interpretations of probability. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2003 edition). http://plato.stanford.edu/archives/sum2003/entries/probability-interpret/. Accessed 14 March 2008.
Halpern, J., Pearl, J. (2005a). Causes and explanations: a structural-model approach. Part I: Causes. *British Journal for the Philosophy of Science*, 56, 843–887.
Halpern, J., Pearl, J. (2005b). Causes and explanations: a structural-model approach. Part II: Explanations. *British Journal for the Philosophy of Science*, 56, 889–911.
Hanson, N. R. (1958). *Patterns of discovery: an inquiry into the conceptual foundations of science*. Cambridge: Cambridge University Press.
Hausman, D. (1983). Are there causal relations among dependent variables? *Philosophy of Science*, 50, 58–81.
Hausman, D. (1993). Linking causal and explanatory asymmetry. *Philosophy of Science*, 60, 435–451.
Hausman, D. (1998). *Causal asymmetries*. Cambridge: Cambridge University Press.
Hausman, D., Woodward, J. (1999). Independence, invariance, and the causal Markov condition. *British Journal for the Philosophy of Science*, 50, 521–583.
Hausman, D., Woodward, J. (2004). Modularity and the causal Markov condition: a restatement. *British Journal for the Philosophy of Science*, 55, 147–161.
Heckman, J. (2005). The scientific model of causality. *Sociological Methodology*, 35(1), 1–97.
Hedström, P., Swedberg, R. (1999a) (Ed.). *Social mechanisms: an analytical approach to social theory*. Cambridge: Cambridge University Press.
Hedström, P., Swedberg, R. (1999b). Social mechanisms: an introductory essay. In P. Hedström, R. Swedberg (Ed.), *Social mechanisms: an analytical approach to social theory* (pp. 1–31). Cambridge: Cambridge University Press.
Hellevik, O. (1984). *Introduction to causal analysis*. London: Allen & Unwin.
Hempel, C. G. (1965). *Aspects of scientific explanation and other essays*. New York: Free Press.
Hempel, C. G., Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2), 135–175. Reprinted in C. G. Hempel, *Aspects of scientific explanation and other essays* (pp. 245–282). New York: Free Press.
Hesslow, G. (1976). Discussion: two notes on the probabilistic approach to causality. *Philosophy of Science*, 43, 290–292.
Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, 98(6), 273–299.
Hitchcock, C. (2004) (Ed.). *Contemporary debates in philosophy of science*. Oxford: Blackwell.
Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–970.
Holland, P. W. (1988). Comment: causal mechanism or causal effect? Which is best for statistical science? *Statistical Science*, 3(2), 186–188.
Holland, P. W., Rubin, D. B. (1988). Causal inference in retrospective studies. *Evaluation Review*, 12(3), 203–231.
Hoover, K. D. (2001). *Causality in macroeconomics*. Cambridge: Cambridge University Press.
Howson, C. (1983). Statistical explanation and statistical support. *Erkenntnis*, 20, 61–78.
Howson, C. (1988). On a recent argument for the impossibility of a statistical explanation of single events. *Erkenntnis*, 29, 113–124.
Howson, C. (2001). The logic of Bayesian probability. In D. Corfield, J. Williamson (Ed.), *Foundations of Bayesianism* (pp. 137–159). Dordrecht: Kluwer.

Howson, C., Urbach, P. (1993). *Scientific reasoning: the Bayesian approach*. Chicago: Open Court. Second edition.

Hume, D. (1739–1740). *A treatise of human nature*. Edited by D. F. Norton, M. J. Norton. Oxford/New York: Oxford University Press, 2000.

Hume, D. (1748). *An enquiry concerning human understanding*. Edited by T. L. Beauchamp. Oxford/New York: Oxford University Press, 1999.

Humphreys, P. (1984). Aleatory explanation expanded. *PSA: Proceedings of the Biannual Meeting of the Philosophy of Science Association*, 1982, vol. 2, 208–223.

Humphreys, P. (1986a). Causality in the social sciences: an overview. *Synthese*, 68, 1–12.

Humphreys, P. (1986b). Quantitative probabilistic causality and structural scientific realism. *PSA*:*Proceedings of the Biannual Meeting of the Philosophy of Science Association*, 1984, vol. 2, 329–342.

Humphreys, P. (1989). *The chances of explanation*. Princeton: Princeton University Press.

Irzik, G. (1986). Causal modelling and the statistical analysis of causation. *PSA: Proceedings of the Biannual Meeting of the Philosophy of Science Association*, 1984, vol. 1: Contributed Papers, 12–23.

Irzik, G. (1996). Can causes be reduced to correlations? *British Journal for the Philosophy of Science*, 47, 249–270.

Irzik, G., Meyer, E. (1987). Causal modeling: new directions for statistical explanation. *Philosophy of Science*, 54, 495–514.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4), 620–630. Reprinted in E. T. Jaynes, *E. T. Jaynes: papers on probability, statistics and statistical physics* (pp. 4–16). Edited by R. G. Rosenkrantz. Dordrecht: Kluwer.

Jaynes, E. T. (1989). *E. T. Jaynes: papers on probability, statistics and statistical physics*. Edited by R. G. Rosenkrantz. Dordrecht: Kluwer.

Jaynes, E. T. (2003). *Probability theory: the logic of science*. Cambridge: Cambridge University Press.

Jeffrey, R. (1966). *The logic of decision*. Chicago: University of Chicago Press. Second edition 1983.

Jeffreys, H. (1939). *Theory of probability*. Reprinted, Oxford: Oxford University Press. Oxford Classics in the Physical Sciences Series. 1998.

Kalleberg, A. L. (1977). Work values and job rewards: a theory of job satisfaction. *American Sociological Review*, 42, 124–143.

Kant, I. (1781). *Critique of pure reason*. Second edition 1787. Translated by N. K. Smith. London: McMillan. 1929.

Kant, I. (1783). *Prolegomena to any future metaphysics*. Translated by P. Gray Lucas. Manchester: Manchester University Press. 1953.

Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan.

Kincaid, H. (1990). Defending laws in the social sciences. *Philosophy of the Social Sciences*, 20, 56–83.

Kincaid, H. (2004). There are laws in the social sciences. In C. Hitchcock (Ed.), *Contemporary debates in philosophy of science* (pp. 168–186). Oxford: Blackwell.

Klein, L. R. (1974). *A textbook of econometrics*. Englewood Cliffs (N.J): Prentice-Hall. Second edition.

Kolmogorov, A. N. (1933). *Foundations of the theory of probability*. New York: Chelsea Publishing. 1950.

Koopman, B. O. (1940). The axioms and algebra of intuitive probability. *Annals of Mathematics*, 41, 269–292.

Korb, K., Wallace, C. (1997). In search of the philosopher's stone: remarks on Humphreys and Freedman's critique of causal discovery. *British Journal for the Philosophy of Science*, 48(4), 543–553.

Kundi, M. (2006). Causality and the interpretation of epidemiological evidence. *Environmental Health Perspectives*, 114, 969–974.

Kyburg, H. E., Smokler, H. E., (1964) (Ed.). *Studies in subjective probability*. New York: Wiley.

Lagiou, P., Adami, H. -O., Trichopoulos, D. (2005). Causality in cancer epidemiology. *European Journal of Epidemiology*, 20, 565–574.
Land, K. C. (1983). Social indicators. *Annual Review of Sociology*, 9, 1–26.
Laplace, P. (1814). *Essai philosophique sur les probabilités*. Paris: Bourgois. 1986.
Larson, J. (1991). *The measurement of health: concepts and indicators*. New York: Greenwood Press.
Laudisa, F. (1999). *Causalità. Storia di un modello di conoscenza*. Roma: Carocci Editore.
Lehnmann, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, 37, 1137–1153.
Lewis, D. (1971). A subjectivist guide to objective chance. Reprinted in D. Lewis, *Philosophical papers Vol. II* (pp. 159–213). Oxford: Oxford University Press. 1986.
Lewis, D. (1973). Causation. Reprinted with postscripts in D. Lewis, *Philosophical Papers Vol. II* (pp. 159–213). Oxford: Oxford University Press.
Lilienfeld, D. E., Stolley, P. D. (1994). *Foundations of epidemiology*. New York: Oxford University Press. Third edition.
Little, D. (1990). *Varieties of social explanations: an introduction to the philosophy of social science*. Boulder: Westview Press.
Little, D. (1993). On the scope and limits of generalizations in the social sciences. *Synthese*, 97, 183–208.
Little, D. (1995a). Causal explanation in the social sciences. *Southern Journal of Philosophy*, 34 (Supplement), 31–56.
Little, D. (1995b) (Ed.). *On the reliability of economic models: essays in the philosophy of economics*. Boston: Kluwer.
Little, D. (1998). *Microfoundations, method and causation. On the philosophy of the social sciences*. New Brunswick, NJ: Transaction Publishers.
Little, D. (2004). Causal mechanisms. In M. S. Lewis-Beck, A. Bryman, T. F. Liao (Ed.), *The Sage Encyclopedia of social science research methods*, Vol. 1. Thousand Oaks, CA: Sage.
Long, J. S. (1983a). *Confirmatory factor analysis*. Beverly Hills: Sage.
Long, J. S. (1983b). *Covariance structure models*. Beverly Hills: Sage.
López-Ríos, O., Mompart, A., Wunsch, G. (1992). Système de soins et mortalité régionale: une analyse causale. *European Journal of Population*, 8(4), 363–379.
Mach, E. (1905). *Knowledge and error*. Dordrecht: Reidel. 1976.
Mackie, J. L. (1974). *The cement of the universe: a study on causation*. Oxford: Clarendon.
Masuy-Stroobant, G. (2002). The determinants of infant mortality: how far are conceptual frameworks really modelled? In R. Franck (Ed.), *The explanatory power of models* (pp. 15–30). Dordrecht: Kluwer.
Masuy-Stroobant, G., Gourbin, C. (1995). Infant health and mortality indicators: their accuracy for monitoring the socio-economic development in the Europe of 1994. *European Journal of Population*, 11(1), 63–84.
Maudlin, T. (2007). *The metaphysics within physics*. Oxford: Clarendon.
McCullag, P. (1989) (Ed.). *Generalized linear models*. London: Chapaman & Hall.
McKim, V., Turner, S. P. (1997) (Ed.). *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences*. Indiana: University of Notre Dame Press.
Meek, C. (1995). Causal inference and causal explanation with background knowledge. In P. Besnard, S. Hanks (Ed.), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 403–410). San Mateo, CA: Morgan Kaufmann.
Mellor, D. H. (1971). *The matter of chance*. London: Cambridge University Press.
Menzies, P., Price, H. (1993). Causation as a secondary quality. *British Journal for the Philosophy of Science*, 44, 187–203.
Mill, J. S. (1843). *A system of logic, ratiocinative and inductive: being a connected view of the principles of evidence and the methods of scientific investigation*. New York: Harper & Brothers. Eighth edition.
Montuschi, E. (2003). *The objects of social science*. London: Continuum.

Montuschi, E. (2006). *Oggettività e scienze umane. Introduzione alla filosofia della ricerca sociale*. Roma: Carocci editore.

Mooney Marini, M., Singer, B. (1988). Causality in the social sciences. *Sociological Methodology*, 18, 347–409.

Morgan, M. (1997). Searching for causal relations in economic statistics. In V. McKim, S. P. Turner (Ed.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 47–80). Indiana: University of Notre Dame Press.

Mosley, W. H., Chen, L. C. (1984). An analytical framework for the study of child survival in developing countries. *Population and Development Review*, 10 (Supplement), 25–45.

Mouchart, M., Russo, F., Wunsch, G. (2008). Structural modelling, exogeneity and causality. In H. Engelhardt, H. -P. Kohler, A. Prskwetz (Ed.), *Causal analysis in population studies: concepts, methods, applications* (Chapter 4). Dordrecht: Springer.

Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw Hill.

Mulaik, S. A. (1985). Exploratory statistics and empiricism. *Philosophy of Science*, 52, 410–430.

Neter, J., Kutner, M. H., Nachtsheim C. J., Wasserman, W. (1996). *Applied linear statistical models*. Chicago: Richard D. Irwin. Fourth edition.

Niiniluoto, I. (1981). Statistical explanation reconsidered. *Synthese*, 48, 437–472.

Norris, P., Inglehart, R. (2003). Islam and the West: testing the 'clash of civilization' thesis. *Comparative Sociology*, 1(3–4), 235–265. http://ksghome.harvard.edu/~pnorris/Acrobat/Clash%20of%20Civilization.pdf. Accessed 14 March 2008.

Otte, R. (1981). A critique of Suppes' theory of probabilistic causality. *Synthese*, 48, 167–189.

Papineau, D. (1985). Causal asymmetry. *British Journal for Philosophy of Science*, 36(3), 273–289.

Papineau, D. (1991). Correlation and causes. *British Journal for Philosophy of Science*, 42(3), 397–412.

Parascandola, M., Weed, D. (2001). Causation in epidemiology. *Journal of Epidemiology and Community Health*, 55, 905–912.

Pearl, J. (1988a). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufman.

Pearl, J. (1988b). Graphs, causality, and structural equation models. *Sociological Methods and Research*, 27(2), 226–284.

Pearl, J. (1990). Jeffrey's rule, passage of experience, and neo-Bayesianism. In H. E. Kyburg et al. (Ed.), *Knowledge representation and defeasible reasoning* (pp. 245–265). Amsterdam: Kluwer.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4), 669–688.

Pearl, J. (2000). *Causality. Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Pearl, J. (2001). Causal inference in statistics: a gentle introduction. Technical Report R-289, Computer Science Department, University of California, Los Angeles.

Pearl, J., Verna, T. S. (1991). A statistical semantics for causation. In *Proceeding, 3rd International Workshop on AI & Statistics*, Fort Lauderdale, FL, 2–5 January 1991. http://bayes.cs.ucla.edu/csl_papers.html. Accessed 14 March 2008.

Pearson, K. (1911). *The grammar of science*. London: A. and C. Black.

Peto, R., Darby, S., Deo, H., Silcocks, P., Whitley, E., Doll, R. (2000). Smoking, smoking cessation, and lung cancer in UK since 1950. *British Medical Journal*, 321, 323–329.

Philips, D., Berman, Y. (2003). Social quality and ethnos communities: concepts and indicators. *Community Development Journal*, 38, 344–357.

Pickett, K., Pearl, M. (2001). Multilevel analysis of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology and Community Health*, 55, 111–122.

Popper, K. R. (1957). The propensity interpretation of the calculus of probability and the quantum theory. In S. Körner (Ed.), *The Colston Papers* (pp. 65–70), vol. 9.

Popper, K. R. (1959). The propensity interpretation of probability. *British Journal for the Philosophy of Science*, 10, 25–42.

Pratt, J. W., Schlaifer, R. (1984). On the nature and discovery of structure. *Journal of the American Statistical Association*, 79(385), 9–21.

Price, H. (1991). Agency and probabilistic causality. *British Journal for the Philosophy of Science*, 42, 157–176.
Price, H. (1992). Agency and causal asymmetry. *Mind*, 101, 501–520.
Price, H. (2004). Models and modals. In D. Gillies (Ed.), *Laws and models in science* (pp. 49–69). London: King's College Publications.
Psillos, S. (2002). *Causation and explanation*. Chesham: Acumen Publishing.
Psillos, S. (2004). A glimpse of the secret connexion: harmonizing mechanisms with counterfactuals. *Perspectives on Science*, 12(3), 288–319.
Psillos, S. (2005). Undetermination. *Encyclopedia of Philosophy*. Gale MacMillan Reference. Second edition. http://www.phs.uoa.gr/~psillos/Publications_files/Underdetermination.doc. Accessed 14 March 2008.
Quetelet, A. (1869). *Physique sociale. Ou essai sur le développement des facultés de l'homme*. Bruxelles: Muquardt.
Ramsey, F. P. (1931). Truth and probability. In F. P. Ramsey, *The foundations of mathematics and other logical essays* (Chapter VII, pp. 156–198). Edited by R. B. Braithwaite. London/London: Kegan Paul/Harcourt, Brace & Co. 1999 electronic edition. http://socserv.mcmaster.ca/econ/ugcm/3ll3/ramseyfp/ramsess.pdf. Accessed 14 March 2008.
Reichenbach, H. (1949). *The theory of probability*. Berkeley: University of California Press.
Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.
Reiss, J. (2001). Natural economic quantities and their measurement. *Measurement in Physics and Economics – Discussion Papers*. Technical Report 14/01, Centre for Philosophy of Natural and Social Science, London School of Economics.
Reiss, J. (2007). Do we need mechanisms in the social sciences? *Philosophy of the Social Sciences*, 37(2), 163–184.
Reiss, J. (2008). *Error in economics: towards a more evidence-based methodology*. London: Routledge.
Roberts, J. T. (2004). There are no laws in the social sciences. In C. Hitchcock (Ed.), *Contemporary debates in the social sciences* (pp. 151–167). Oxford: Balckwell.
Robinson, W. (1950). Ecological correlations and the behaviour of individuals. *American Sociological Review*, 15, 351–357.
Rosen, D. (1978). In defence of a probabilistic theory of causality. *Philosophy of Science*, 45, 604–613.
Rosenbaum, P. R. (1984). From association to causation in observational studies. The role of tests of strongly ignorable treatment assignment. *Journal of the American Statistical Association*, 385, 41–48.
Rosenbaum, P. R., Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
Rubin, D. (1974). Estimating causal effects of treatments in randomized and non randomized studies. *Journal of Educational Psychology*, 66(5), 688–701.
Rubin, D. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, 6, 34–58.
Russell, B. (1912–1913). On the notion of cause. *Proceedings of the Aristotelian Society*, 13, 1–26.
Russo, F. (2006a). Salmon and van Fraassen on the existence of unobservable entities: a matter of interpretation of probability. *Foundations of Science*, 11(3), 221–247.
Russo, F. (2006b). The rationale of variation in methodological and evidential pluralism. *Philosophica*, 77, 97–124.
Russo, F. (2007). Frequency-driven probabilities in quantitative causal analysis. *Philosophical Writings*, 32, 32–56.
Russo, F., Mouchart, M., Ghins, M., Wunsch, G. (2006). Causality, structural modelling and exogeneity. Discussion Paper 0601, Institut de Statistique, Université catholique de Louvain, Belgium.
Russo, F., Williamson J. (2007a) (Ed.). *Causality and probability in the sciences*. Texts in Philosophy Series. London: College Publications.

Russo, F., Williamson, J. (2007b). Interpreting probability in causal models for cancer. In F. Russo, J. Williamson (Ed.), *Causality and probability in the sciences* (pp. 217–242). Texts in Philosophy Series. London: College Publications.
Russo, F., Williamson, J. (2007c). Interpreting causality in the health sciences. *International Studies in Philosophy of Science*, 21(2), 157–170.
Salmon, W. C. (1967). *Foundations of scientific inference*. Pittsburgh: University of Pittsburgh Press.
Salmon, W. C. (1971). Statistical explanation. In W. C. Salmon et al. (Ed.), *Statistical explanation and statistical relevance* (pp. 29–87). Pittsburgh: University of Pittsburgh Press.
Salmon, W. C. (1977). Objectively homogeneous references classes. *Synthese*, 36, 399–414.
Salmon, W. C. (1980). Probabilistic causality. *Pacific Philosophical Quarterly*, 61, 50–74.
Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
Salmon, W. C. (1988). Dynamic rationality. In J. Fetzer J. (Ed.), *Probability and causality* (pp. 3–42). Dordrecht: Reidel.
Salmon, W. C. (1990). Causal propensities: statistical causality vs. aleatory causality. *Topoi*, 9, 95–100.
Salmon, W. C. (1998). *Causality and explanation*. New York: Oxford University Press.
Salmon, W. C. et al. (1971). *Statistical explanation and statistical relevance*. Pittsburgh: University of Pittsburgh Press.
Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
Simon, H. (1952). On the definition of the causal relation. *Journal of Philosophy*, 49(16), 517–528.
Simon, H. (1953). Causal ordering and identifiability. In W. C. Hood, T. C. Koopmans (Ed.), *Studies in econometric method* (pp. 49–74). New York: Wiley.
Simon, H. (1954). Spurious correlation: a causal interpretation. *Journal of the American Statistical Association*, 49(267), 467–479.
Simon, H. (1979). The meaning of causal ordering. In K. R. Merto, J. J. Coleman, P. H. Rossi (Ed.), *Qualitative and quantitative social research* (pp. 65–81). New York: Free Press.
Skyrms, B. (1988). Probability and causation. *Journal of Econometrics*, 39, 53–68.
Snijders, T. A. B., Bosker, R. J. (2004). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London: Sage. Fourth edition.
Sober, E. (1984). Two concepts of cause. *PSA: Proceedings of the Biannual Meeting of the Philosophy of Science Association*, 1982, vol. 2, 405–424.
Sober, E. (1986). Causal factors, causal influence, causal explanation. *Proceedings of Aristotelian Society*, 60, 97–136.
Spirtes, P., Glymour, C., Scheines, R. (1991). From probability to causality. *Philosophical Studies*, 91, 1–36.
Spirtes, P., Glymour, C., Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer.
Spirtes, P., Richardson, T., Meek, C., Scheines, R., Glymour, C. (1998). Using path diagrams as a structural equation modeling tool. *Sociological Method and Research*, 27(2), 182–225.
Steuer, M. (2003). *The scientific study of society*. Boston: Kluwer.
Stone, R. (1993). The assumptions on which causal inferences rest. *Journal of the American Statistical Association*, 55(2), 455–466.
Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North Holland.
Suppes, P. (1982). Problems of causal analysis in the social sciences. *Epistemologia*, Special Issue 1982, 239–250.
Suppes, P. (2002). *Representation and invariance of scientific structures*. Stanford: CSLI Publications.
Susser, M. (2001). Glossary: causality in public health sciences. *Journal of Epidemiology and Community Health*, 55, 376–378.
Susser, M. W. (1973). *Causal thinking in the health sciences*. New York: Oxford University Press.

Susser, M., Susser, E. (1996). Choosing a future for epidemiology ii: from black box to Chinese box and ecoepidemiology. *American Journal of Public Health*, 86, 674–677.

Thompson, P. (2006). Bayes p-values. In S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic (Ed.), *Encyclopedia of Statistical Sciences*. Hoboken, NJ: Wiley-Interscience. Second edition.

Toulemon, L. (2006). Les modèles de régression. In G. Caselli, J. Vallin, G. Wunsch (Ed.), *Démographie: analyse et synthèse* (pp. 359–374). Volume VIII Observation, méthodes auxiliaires, enseignement et recherche. Paris: Editions de l'Institut National d'Etudes Démographiques.

Urbach, P. (1989). Random sampling and the principles of estimation. *Proceedings of the Aristotelian Society*, 89, 143–164.

Vandresse, M. (2005). Characteristics of the newborn at birth and health status: a congested two-frontiers approach. Unpublished manuscript.

Vandresse, M. (2008). Late fertility: its causal effects on health of the newborn and its implications in fertility decision process. PhD Thesis. Université catholique de Louvain.

van Bouwel, J. (2004). Individualism and holism, reduction and pluralism: a comment on Keith Sawyer and Julie Zahle. *Philosophy of the Social Sciences*, 34(4), 527–535.

van Fraassen, B. C. (1980). *The scientific image*. Oxford: Clarendon Press.

van Fraassen, B. C. (1983). Calibration: a frequency justification for personal probabilities. In R. S. Cohen, L. Laudan (Ed.), *Physics, philosophy and psychoanalysis: essays in honour of Adolph Grümbaum* (pp. 295–319). Dordrecht: Reidel.

van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford: Clarendon.

Venn, J. (1876). *The logic of chance: an essay on the foundations and province of the theory of probability*. London: Macmillan. Second edition.

Vineis, P. (2003). Causality in epidemiology. *Sozial und Praventiv Medizin*, 48, 80–87.

von Bertalanffy, L. (1969). *General system theory: foundations, development, applications*. New York: Braziller.

von Mises, R. (1957). *Probability, statistics and truth*. New York: Macmillan. Second edition.

von Wright, G. (1971). *Explanation and understanding*. Ithaca: Cornell University Press.

Wackerly, D. D., Mendenhall W., Scheaffer R. L. (2002). *Mathematical statistics with applications*. Pacific Grove: Duxbury. Sixth edition.

Williams, M., Williamson, J. (2006). Combining argumentation and Bayesian nets for breast cancer prognosis. *Journal of Logic, Language and Information*, 15, 155–178.

Williamson, J. (2005a). *Bayesian nets and causality. Philosophical and computational foundations*. Oxford: Oxford University Press.

Williamson, J. (2005b). Causality. In D. Gabbay, F. Guenthner (Ed.), *Handbook of philosophical logic* (pp. 131–162). volume 13. Springer.

Williamson, J. (2005c). Philosophies of probability: objective Bayesianism and its challenges. In A. Irvine (Ed.), *Handbook of the philosophy of mathematics*. Volume four of *Handbook of Philosophy of Science*. Elsevier. http://www.kent.ac.uk/secl/philosophy/jw/2004/philprob.pdf. Accessed 17 March 2008.

Williamson, J. (2007). Motivating objective Bayesianism: from empirical constraints to objective probabilities. In W. L. Harper, G. R. Wheeler (Ed.), *Probability and inference: essays in honour of Henry E. Kyburg Jr* (pp. 155–183). London: College Publications.

Woodward, J. (1993). Book review: Humphreys, P., 'The chances of explanation: causal explanation in social, medical and physical sciences'. *Philosophy of Science*, 60(4), 671–673.

Woodward, J. (1999). Causal interpretation in systems of equations. *Synthese*, 121, 199–247.

Woodward, J. (2003). *Making things happen: a theory of causal explanation*. Oxford: Oxford University Press.

Wonnacott, T. H., Wonnacott, R. J. (1990). *Introductory Statistics*. New York: Wiley. Fifth edition.

Worrall, J. (2002). What evidence in evidence-based medicine? *Philosophy of Science*, 69, 316–330.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557–585.

Wright, S. (1934). The method of path coefficient. *Annals of Mathematical Statistics*, 5(3), 161–215.

Wunsch, G. (1984). Theories, models and knowledge: the logic of demographic discovery. *Genus*, 40(1–2), 1–18.
Wunsch, G. (1986). Causal connections in causal modelling. *Genus*, 42(3–4), 1–12.
Wunsch, G. (1988). *Causal theory and causal modelling*. Leuven: Leuven University Press.
Wunsch, G. (1995). God has chosen to give the easy case to the physicists. In *Evolution or revolution in European population. European Population Conference. 1 Plenary Session* (pp. 201–224). Milano: Franco Angeli.
Wunsch, G. (2007). Confounding and control. *Demographic research*, 6(4), 95–120.
Yule, G. U. (1897). On the theory of correlation. *Journal of Royal Statistical Society*, 60, 812–854.
Zellner, A. (1988). Causality and causal laws in economics. *Journal of Econometrics*, 39, 7–21.
Zeller, R. A., Carmines, E. G. (1980). *Measurement in the social science. The link between theory and data*. Cambridge: Cambridge University Press.

Index

A
Agent probability, 201
Alternative hypothesis, *see* Hypothesis testing
Aristotle, 6
Assumptions of causal models
  about errors, 59
  as constraints, 125
  homoscedasticity, 59
  independence and distributions of errors, 67
  linear dependence, 67
  linearity and normality, 58
  no-measurement error, 59
  non-correlation between errors, 61
  non-correlation between unobserved variables and errors, 61
  non-correlation between explanatory variables and errors, 61
  separability and additivity, 58
  stability or invariance, 59, 204
  vs. criteria to assess causality, 120
At-at theory of mark transmission, 194
Average causal effect, 63, 98
Average man, 101

B
Background knowledge, 31, 76, 77, 121, 125, 131, 164, 167, 170, 206
  vs. foundations, 122
Bayes' theorem, 135, 143, 181
  and likelihood, 183
Bayesian networks, 56
Black box, 28, 125, 159
Blalock, H. M., 78, 82, 93
Boudon, R., 67, 85
Bunge, M., 93, 159, 160, 161, 188

C
Calibration, 144
Capacity, 105, 157, 161
Carnap, R., 136, 142, 143, 147
Cartwright, N., 47, 71, 105, 116, 128, 157, 161
Case studies
  farmers' migration, 24, 65, 190, 191
  health and wealth, 23, 24, 114, 115, 126, 127, 128
  health system and mortality, 120
  Islamic culture and gender equality, 110, 111, 112
  job satisfaction, 25, 167, 168, 169, 170
  mother's education and child survival, 21–23, 27–29, 110, 114, 162, 164
  newborn's characteristics and health status at birth, 98, 99, 100
  obesity and educational level, 118
  smoking and lung cancer, 20, 21, 110, 115, 121
Causal asymmetry, 57, 119
Causal belief, 206
Causal chain, 32, 37, 43, 85, 156, 157, 158
Causal context, 29, 30, 31, 70, 76, 77, 119, 125
Causal directionality, 83
Causal Markov condition, 95, 128, 129, 132, 206
Causal modelling
  as a model of explanation, 166
  as modelling mechanisms, 161
  assumptions, *see* Assumptions of causal models
  contingency tables, 67–70, 97–98
  covariance structure models, 60–61, 71, 97

Causal modelling (*cont.*)
difficulties and weaknesses, 75–90
directed acyclic graphs, 56, 128
double-frontier approach, 98–101
Granger-causality, 23, 61, 62, 63, 126, 127
multilevel analysis, 64, 65, 66, 67, 188–190
path models and causal diagrams, 54–60, 96–97
Rubin's model, 63, 64
theorising vs. modelling, 77
Causal models
vs. associational models, 116–132
Causal ordering, 57, 83, 119, 127
Causal process, 17–19, 26, 28, 157, 194–197
vs. pseudo-process, 194
Causal realism, 12
vs. objectivity, 12
Causal relation
among quantitative properties, 50, 51, 52
direction of, 43
epistemic access to, 29, 94
strength of, 180
Causal structures, 85
Causality
agency theory, 200–203
aleatory, 17, 18, 26, 49
as mental construct, 205
as pattern of knowledge, 5–8
counterfactual account, 197, 200
epistemic theory, 205–207
generic vs. single-case, 49, 134, 135, 174, 177–179
indeterministic, 17
manipulability theory, 106, 203–205
mechanist account, 33, 194–197
mind-dependent, 6
population-level vs. individual level, 39, 49, 179–183
probabilistic, 10, 17, 36
probabilistic theories of, 36, 52, 95, 96
probabilistic vs. deterministic, 44, 45, 46
production, 93, 94
rationale of, 2, 10, 30, 92–114, 211
rationale vs. definition, 93
regularity theory, 9, 32, 107, 208
singularist account, 207–209
statistical, 18, 26–30
token, 173–177
twofold, 46–49
type-level vs. token-level, 49
Causation as a secondary quality, 200
Cause
contributing, 105
negative, 42, 46, 49
prima facie, 40
spurious, 41
sufficient, 45
temporal priority of, 37, 43, 51
type vs. token, 46, 47, 48
Causes of effects vs. effects of causes, 83
Clash of civilization thesis, 111
Coefficient of determination, 59, 96, 116
Conditionalisation 137, 149
Collingwood, R., 201
Conditional independence, 62
Confirmatory logic, 71
Conjunctive fork, 17
Connecting principle, 179–183
Constant conjunction, 6, 17, 107, 109, 198
Correlation coefficient, 8, 50, 57, 123, 181
Counterfactual, 28, 83
Courgeau, D., 24, 65, 87, 149, 151, 188, 190
Covariance, 50, 51
Covariation, 60–61, 88–89, 107
Covering law model, 73
Cox, D. T., 84

D
Data snooping, 87
Dawid, A. P., 83
Decision theory, 84, 147, 201
de Finetti, B., 138, 148, 149
Dependence, 51
Dowe, P., 194–197
Ducasse, J. C., 207–209
Dupré, J., 157
Durkheim, E., 102, 104, 131, 190
Dutch book arguments, 138, 150

E
Edwards, A. W. F., 181
Eells, E., 173–177
Epistemology of causality, 2, 25, 26, 30, 91–132, 211
Equations, 54–56
Errors, 56
Evidence, 13, 29, 31, 70, 78, 84, 120, 179
Examples
contraceptive pills, pregnancy, and thrombosis, 44, 95
decisions in the Norwegian court, 68
falling barometer, 42, 51
fertility study, 57
golf ball and squirrel's kick, 47, 173–177

Holmes, Moriarty, and Watson, 39
McCulloch and Pitt's model of the neuron, 158
particle transition, 195
price and demand equilibrium, 54, 115
season, rain, and sprinkler, 129
socio-economic status, asbestos, smoking, and lung cancer, 132
sprayed plant, 47
Exchangeability, 148–150
Exogeneity, 60
Experimentation, 94
Explanation
and mechanisms, 156–172
covering-law model of, 73, 112
epistemic control, 170
goodness of, 170
hypothetico-deductive, 170
metaphysical control, 170
S-R model of, 97–98, 110–112
statistical control, 170
through mechanisms, 165–171
why-question, 113
Exploratory data analysis, 74, 75

F
Faithfulness, 123
Fallacies
atomistic and ecological, 65, 185
Features of causal models
causal assumptions, 123
causal context, 119
conceptual hypothesis, 119
extra-statistical assumptions, 119
statistical assumptions, 117–118, 123
Fisher, R. A., 67, 130, 131, 181
Foundations of epidemiology, 104, 105
Franck, R., 85, 158
Freedman, D. A., 116, 151

G
Gérard, H., 77
Gasking, D., 201
Generalisation, 7, 16, 73, 90, 94, 173, 186, 204
change-relating, 106, 204
empirical, 28, 73, 80, 161
universal, 28, 80
Gillies, D., 140, 145, 148, 149
Glymour, C., 74
Goldstein, H., 65
Good, I. J., 36–40, 74
Granger, C. W. J., 23, 61

H
Haavelmo, T., 56, 96, 130
Hacking, I., 140, 182
Hausman, D., 105, 116, 203–205
Heckman, J., 84
Hempel, C. G., 73
Hitchcock, C., 80
Holism vs. individualism, 189
Holland, P. W., 63, 116
Hume, D., 6, 93, 107, 198
Humphreys, P., 105, 130
Hypothesis, 10, 22, 32, 60, 77
conceptual, 32, 70, 125
Hypothesis testing, 86, 94, 151
Hypothetico-deductive methodology, 70–75, 125, 167–170, 205
deductive vs. inductive inference, 72
vs. inductive methodology, 72, 74, 75

I
Indicator, 27, 81, 82
Individual
statistical vs. real, 189

J
Jaynes, E. T., 138, 152
Jeffreys, H., 136

K
Kant, I., 7
Keynes, J. M., 136
Koopmans, T. C., 56

L
Laplace, P., 136
Laws of nature, 73, 113
variational reading of, 113
Lazarfeld, P., 67, 85
Levels of analysis, 183–191
by aggregation, 187
by discipline, 187
Levels of causation, 45, 172–183
vs. levels of analysis, 186, 187
Lewis, D., 145, 197–200
Likelihood and support, 181
Little, D., 28, 156

M
Mach, E., 8
Mackie, J. L., 85
Maudlin, T., 113
Measurement, 50, 59–61, 81, 82, 170

Mechanism, 12, 22, 25, 26, 28, 29, 32, 33, 80, 83, 89, 93, 113, 132, 156–172
and explanation, 156–172
generative, 158, 159
lawful, 161
material structure vs. functional structure, 159
mixed, 163–165
physical, 16, 160
vs. causal chain, 85, 158
vs. decision-making process, 171, 172
vs. system, 160
Menzies, P., 200–203
Metaphysics of causality, 2, 26, 93, 173–179, 212
Methodology of causality, 2, 19, 25, 26, 30, 53–90
Mill's methods of experimetal inquiry, 102
Mill, J. S., 7, 101, 102
Modal realism, 198
Modularity, 130, 157
Montuschi, E., 79
Mulaik, S., 74
Multilevel analysis, *see* Causal modelling, multilevel analysis
Multilevel modelling, *see* Causal modelling, multilevel analysis

N
Null hypothesis, *see* Hypothesis testing

O
Objectivity
epistemic vs. ontic, 13

P
p-value, 86, 87, 152
Parameter, 55–57
Path analysis, *see* Causal modelling, path models and causal diagrams
Path coefficient, 56
Pearl, J., 74, 107, 116, 199, 200
Pearson, K., 8, 67
Popper, K., 70, 137, 145
Possible-world semantics, 197
Potential outcome, 63
Price, H., 200–203
Principal principle, 145, 146
Probabilistic theories of causality, *see* Causality, probabilistic theories of
Probability
as a twofold concept, 140, 141
as guide for action, 153, 154
Bayesian interpretations, 137–139
classical and logical interpretation, 135, 136
credence-driven, 145, 146
empirically-based Bayesian interpretation, 138
empirically-based vs. objective Bayesianism, 150–154
frequency-driven, 139–150
frequency-driven vs. credence-driven, 140
frequentist interpretation, 136
Kolmogorov's aziomatisation, 135
objective Bayesian interpretation, 138
physical, 37
physical interpretation, 136, 137
propensity interpretation, 136
subjective, 37
subjective Bayesian interpretation, 138
Probability raising requirement, 45–47, 109, 173–177, 194
Propensity, *see* Probability, propensity interpretation
Psillos, S., 84, 208

Q
Quetelet, A., 101

R
Ramsey, F. P., 138–142
Rational decision-making, 147
Rationale of causality, *see* Causality, rationale of
Rationale of variation, *see* Variation, rationale of
Regular succession, 6, 96, 108, 198, 208, 215
Regularity
accidental, 198
governing, 28
phenomenal, 28
statistical, 17
universal, 80
vs. invariance, 124
Reichenbach, H., 17, 137, 194
Reiss, J., 81
Relative risk, 21, 31, 110
Robinson, W., 65, 185
Rubin, D., 63
Russell, B., 8

S
Salmon, W. C., 17–20, 40, 112, 137, 138, 140, 143, 144, 157, 194–197
Screening-off, 17, 40, 103

Significance level, *see* Hypothesis testing
Sober, E., 179–183
Social quality framework, 81
Spirtes, P., 76
Statistical relevance, 17, 36–38, 40, 43
Stone, R., 116
Structural equations models, *see* Causal modelling, path models and causal diagrams
Suppes, P., 40–43, 131, 157

T
Test statistic, *see* Hypothesis testing
Theory ladeness, 81
Treatment and control group, 63
Type I and type II error, 86, 152, 153

U
Underdetermination, 84

V
Validity, 87–90
  construct, 88
  external, 88
  internal, 88
  statistical, 88
van Fraassen, B. C., 112, 144, 145
Variable, 26, 49, 184–185
  categorical, 67, 184
  nominal, 67, 184
  ordinal, 67, 184
  types of, 80, 184, 185
Variation, 3–5, 11, 52, 80, 87, 89, 92–132, 152, 161, 162, 168, 172, 189–192, 193–210
  vs. invariance, 93
  as a precondition, 108
  in agency theory, 200–203
  in causes of states, 112–114
  in counterfactual approaches, 197–200
  in epistemic theory, 205–207
  in homogenous populations, 109–112
  in manipulability theory, 203–205
  in mechanisms, 162
  in mechanist approaches, 194–197
  in single instances, 207–209
  objections to, 107–114
  rationale of, 92–114
  taxonomy of, 114–116
  vs. dependence, 107
  vs. invariance, 30, 89, 105, 128, 204
  vs. regularity, 28, 30, 89, 93, 107, 109
von Mises, R., 137
von Wright, G., 201

W
Williamson, J., 76, 139, 151, 171, 205, 207
Woodward, J., 105, 106, 157, 203–205
Wright, S., 19, 56, 58
Wunsch, G., 78, 79, 109–112

Y
Yule, U., 67