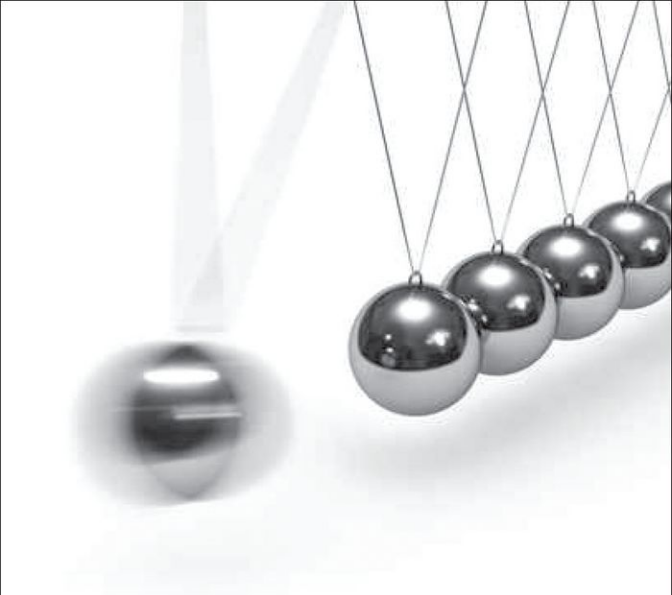

Causality in the Sciences

Phyllis McKay Illari | Federica Russo | Jon Williamson

OXFORD

Causality in the Sciences

This page intentionally left blank

Causality in the Sciences

Edited by

Phyllis McKay Illari
Research Fellow, University of Kent

Federica Russo
Research Associate, University of Kent

Jon Williamson
Professor of Reasoning, Inference and Scientific Method, University of Kent

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trademark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First published 2011

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
Library of Congress Control Number: 2011922687
Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
CPI Antony Rowe, Chippenham, Wiltshire
ISBN 978–0–19–957413–1
1 3 5 7 9 10 8 6 4 2

Contents

List of Contributors

Ahn, Woo-kyoung Department of Psychology, Yale University, woo-kyoung.ahn@yale.edu

Aizawa, Ken Department of Philosophy, Centenary College of Louisiana, ken.aizawa@gmail.com

Aliferis, Constantin Center for Health Informatics and Bioinformatics, New York University, constantin.aliferis@nyumc.org

Baetu, Tudor M. ARHU-Philosophy, University of Maryland, tbaetu@hotmail.com

Bogen, Jim HPS Department, University of Pittsburgh, rtjbog@comcast.net

Boniolo, Giovanni FOM, Firc Institute of Molecular Oncology & Dipartimento di Medicina, Chirurgia e Odontoiatria, Università di Milano, giovanni.boniolo@ifom-ieo-campus.it

Broadbent, Alex Department of Philosophy, University of Johannesburg, a.b.broadbent@gmail.com

Cartwright, Nancy Department of Philosophy, Logic and Scientific Method, London School of Economics, and Department of Philosophy, University of California, San Diego, N.L.Cartwright@lse.ac.uk

Cooper, Gregory Department of Biomedical Informatics, University of Pittsburgh, gfc@pitt.edu

D'Abramo, Flavio Department of Philosophical and Epistemological Studies, University of Rome 'La Sapienza', and Department of Epidemiology and Public Health, Imperial College London, flavio.dabramo@gmail.com

Dawid, A. Philip Centre for Mathematical Sciences, University of Cambridge, A.P.Dawid@statslab.cam.ac.uk

Dowe, Phil Philosophy, University of Queensland, p.dowe@uq.edu.au

Elisseeff, André Nhumi Technologies, Zurich, andre@nhumi.com

Faraldo, Rossella Department of Physics, University of Padova, and INFN Padova Section, faraldo@pd.infn.it

Fennell, Damien Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, D.J.Fennell@lse.ac.uk

Fitelson, Branden Department of Philosophy, Rutgers University, branden@fitelson.org

Geneletti, Sara Department of Statistics, London School of Economics and Political Science, S.Geneletti@lse.ac.uk

Gillett, Carl Department of Philosophy, Northern Illinois University, carl.gillett@gmail.com

Gillies, Donald Department of Science and Technology Studies, University College London, donald.gillies@ucl.ac.uk

Glennan, Stuart Department of Philosophy and Religion, Butler University, sglennan@butler.edu

Groff, Ruth Department of Political Science, Saint Louis University, rgroff@slu.edu

Guyon, Isabelle Clopinet, Berkeley, California, guyon@clopinet.com

Gyenis, Balázs Department of History and Philosophy of Science, University of Pittsburgh, gyepi@hps.elte.hu

Hassani, Hossein School of Mathematics, Cardiff University, HassaniH@cf.ac.uk

Hitchcock, Christopher Division of Humanities and Social Sciences, California Institute of Technology, cricky@its.caltech.edu

Holmes, Dawn E. Department of Statistics and Applied Probability, University of California, holmes@pstat.ucsb.edu

Hope, Lucas Bayesian Intelligence, 2/21 The Parade, Clarinda, VIC 3169, Australia, lucas.r.hope@gmail.com

Hoover, Kevin D. Department of Economics and Department of Philosophy, Duke University, kd.hoover@duke.edu

Illari, Phyllis McKay Philosophy, School of European Culture and Languages, University of Kent, p.mckay@kent.ac.uk

Khan, Aneire MRC/HPA Centre for Environment and Health, School of Public Health, Imperial College London, aneire.khan@imperial.ac.uk

Kincaid, Harold Departments of Philosophy and Epidemiology, University of Alabama at Birmingham, hkincaid@me.com

Kleinberg, Samantha Computer Science Department, New York University, samantha@cs.nyu.edu

Korb, Kevin B. School of Info Tech, Monash University, kbkorb@gmail.com

Kuhlmann, Meinard Institute of Philosophy, University of Bremen, meik@uni-bremen.de

Lagnado, David Division of Psychology and Language Sciences, University College London, d.lagnado@ucl.ac.uk

Lappi, Otto Cognitive Science, Institute of Behavioural Sciences, University of Helsinki, olappi@mappi.helsinki.fi

Lemeire, Jan ETRO Department, Vrije Universiteit Brussel, jlemeire@etro.vub.ac.be

Leuridan, Bert Centre for Logic and Philosophy of Science, Ghent University, Bert.Leuridan@UGent.be

Little, Daniel Philosophy, University of Michigan-Dearborn, dlittle30@gmail.com

Luhmann, Christian Department of Psychology, Stony Brook University, christian.luhmann@stonybrook.edu

Machamer, Peter History and Philosophy of Science, University of Pittsburgh, pkmach@pitt.edu

Markus, Keith A. John Jay College of Criminal Justice of The City University of New York, kmarkus@aol.com

Millstein, Roberta L. Department of Philosophy, University of California, Davis, RLMillstein@ucdavis.edu

Mishra, Bud Computer Science Department, New York University, mishra@nyu.edu

Mouchart, Michel Institut de statistique, biostatistique et sciences actuarielles (ISBA), Université catholique de Louvain, Belgium, Michel.Mouchart@uclouvain.be

Nyberg, Erik P. History and Philosophy of Science Program, The University of Melbourne, e.nyberg@pgrad.unimelb.edu.au

Patterson, Kerry Professor of Econometrics, School of Economics, University of Reading, k.d.patterson@reading.ac.uk

Pearl, Judea Cognitive Systems Lab, Computer Science Department, University of California, judea@cs.ucla.edu

Pellet, Jean-Philippe IBM Zurich Research Labs, Zurich, jppellet@gmail.com

Pendergraft, Garrett Philosophy Department, Pepperdine University, garrett.pendergraft@pepperdine.edu

Psillos, Stathis Department of Philosophy and History of Science, University of Athens, psillos@phs.uoa.gr

Rédei, Miklós Department of Philosophy, Logic and Scientific Method, London School of Economics, M.Redei@lse.ac.uk

Reiss, Julian Faculty of Philosophy, Erasmus University Rotterdam, reiss@fwb.eur.nl

Rottman, Benjamin Psychology, Yale University, benjamin.rottman@yale.edu

Rusanen, Anna-Mari Philosophy of Science Group, Department of Philosophy, History, Culture and Arts Studies, University of Helsinki, anna-mari.rusanen@helsinki.fi

Russo, Federica Philosophy, School of European Culture and Languages, University of Kent, f.russo@kent.ac.uk

Saggion, Antonio Department of Physics, University of Padova, saggion@pd.infn.it; antonio.saggion@unipd.it

Sarkar, Sahotra Department of Philosophy and Section of Integrative Biology, University of Texas at Austin, sarkar@mail.utexas.edu

Silva, Ricardo Department of Statistical Science, University College London, ricardo@stats.ucl.ac.uk

Sloman, Steven A. Cognitive, Linguistic, and Psychological Sciences, Brown University, steven_sloman@brown.edu

Soofi, Abdol S. Department of Economics, University of Wisconsin-Platteville, Soofi@uwplatt.edu

Spirtes, Peter Department of Philosophy, Carnegie Mellon University, ps7z@andrew.cmu.edu

Statnikov, Alexander Center for Health Informatics and Bioinformatics, New York University, Alexander.Statnikov@med.nyu.edu

Steenhaut, Kris ETRO Department, Vrije Universiteit Brussel, kris.steenhaut@vub.ac.be

Thompson, R. Paul Institute for the History and Philosophy of Science and Technology, and Department of Ecology and Evolutionary Biology, University of Toronto, p.thompson@utoronto.ca

Touhafi, Abdellah ETRO Department, Vrije Universiteit Brussel, abdellah.touhafi@vub.ac.be

Vineis, Paolo MRC/HPA Centre for Environment and Health, School of Public Health, Imperial College London, p.vineis@imperial.ac.uk

Walsh, Clare R. School of Psychology, University of Plymouth, clare.walsh@plymouth.ac.uk

Weber, Erik Centre for Logic and Philosophy of Science, Ghent University, Erik.Weber@UGent.be

Williamson, Jon Philosophy, School of European Culture and Languages, University of Kent, j.williamson@kent.ac.uk

Zhigljavsky, Anatoly School of Mathematics, Cardiff University, ZhigljavskyAA@cf.ac.uk

This page intentionally left blank

PART I
Introduction

This page intentionally left blank

·1·
Why look at causality in the sciences? A manifesto

Phyllis McKay Illari, Federica Russo, and Jon Williamson

Abstract

This introduction to the volume begins with a manifesto that puts forward two theses: first, that the sciences are the best place to turn in order to understand causality; second, that scientifically-informed philosophical investigation can bring something to the sciences too. Next, the chapter goes through the various parts of the volume, drawing out relevant background and themes of the chapters in those parts. Finally, the chapter discusses the progeny of the papers and identifies some next steps for research into causality in the sciences.

1.1 A manifesto

One might think that the sciences are the last place one should look to gain insights about causality. This is because, due to influential arguments of Karl Pearson, Ernst Mach and Bertrand Russell at the turn of the twentieth century, research scientists have for a long time taken great pains to eradicate causal talk from their research papers and to talk instead of associations, correlations, risk factors and other ephemeral properties of data. Thus the traditional home of the study of causality has been within the field of metaphysics in philosophy – a field that has in its turn been treated sceptically by many scientists.

Our first thesis is that, on the contrary, the sciences are the best place to turn in order to understand causality. We maintain this thesis for a variety of reasons.

First, as explained in Section 1.2.5, causal talk became more respectable in the sciences at the turn of the twenty-first century, thanks to attempts to mathematize the notion of cause. It is now becoming clearer that causal reasoning is of central concern to scientists in many fields, as well as to philosophers, and it is fruitfully pursued as a project of mutual concern.

Second, although causal talk was unfashionable in the twentieth century, causality never really went away: scientists' claims were always intended

to inform policy, experiment and technology, and such applications require causation, rather than mere association which tells us nothing about what happens when we intervene to change the world.

Third, the concept of cause is changing, and the sciences are at the forefront of these changes. In Aristotle's time causality was understood as explanation in general: the search for causes was a search for 'first principles', which were meant to be explanatory. However, now causal explanation is usually thought of as just one kind of explanation. In the modern era, causality became tied up with the notion of determinism, the prevailing scientific view of the world in Newtonian times. But determinism fell out of favour in science due to the advent of quantum mechanics. Moreover, a non-deterministic notion of cause became increasingly relevant to science (in medicine, for example, claims like 'smoking causes cancer', where the cause is not sufficient to ensure the effect, became quite acceptable), and causality lost its deterministic connotations. If attempts within science to mathematize the notion of cause should prove successful (though this is controversial), the current concept of cause may be replaced by some formal explication, as happened so systematically with the concept of probability. It is science that is driving change in the concept of cause.

Fourth, the field of metaphysics generally benefits immeasurably from interactions with the sciences. Our understanding of time and space, for example, is derived from the use of these notions in physics, just as our understanding of what an organism is (and could be) is derived from the biological study of organisms. It is part of the job of any scientific field to decide what the constituents of its field are, whether that is four-dimensional space-time, bacteria, or market transactions. This is the same question that is faced at a higher degree of abstraction by the metaphysician concerned with what the constituents of the world are. It is bizarre to try to answer those questions without looking at how the same questions are dealt with in the sciences.

Our second thesis is that scientifically-informed philosophical investigation can bring something valuable to the sciences, too. As can be seen in this volume, many scientific fields are wrestling with the same methodological problems concerning causality. Different sciences use different languages and different paradigmatic examples, which can obscure the fact that they are facing the very same problems. But philosophers of science are in a natural position to identify common ground in the methods they encounter across the sciences. These philosophers are becoming increasingly well-informed about the sciences and so able to exploit that position in order to identify best practice. Of course philosophers are also well placed to identify any conceptual problems that they encounter in the methods developed in the sciences and to clarify the very concept of cause that these methods appeal to.

We think, then, that the most promising way forward in understanding causality and making methodological progress is as a mutual project between philosophically-minded scientists and scientifically-informed philosophers. We hope that this volume is testimony to the fruitfulness of this way of looking at causality in the sciences.

1.2 The core issues

1.2.1 Health sciences

While biomedical issues have long been a concern of ethicists and phenomenologists, only very recently have the health sciences become prominent in the debates of philosophers of science and philosophers of causality. It is now clear that the health sciences are an inspiring source for methodological, epistemological and metaphysical issues concerning causation. The chapters in this part of the volume testify to the increasing awareness of both philosophers and practising scientists that biomedical research shares with other domains a number of concerns, from the conditions for inferring causation from correlational data to the definition, use, and role of mechanisms. What triggered philosophers to pay more attention to this domain has been the rise of the so-called *evidence-based medicine* (EBM) movement. Although the first works by the epidemiologist Archie Cochrane going in this direction date from the early 1970s, the term was coined and started to be customarily used only in the early 1990s. The main result has been the production of the so-called 'evidence-hierarchy', i.e. a list ranking methods for causal inference from the strongest (notably, meta-analyses of randomized controlled trials) to the weakest (notably, expert opinion). *Evidence*, it seems, is the pillar of science and the tenets of EBM are well-entrenched. But these strongholds have been under attack for the last 10 years at least. The battle to set the debate straight is happening in this volume too.

For instance, in *Causality, theories, and medicine* Paul Thompson argues against RCTs as the gold standard of causal inference in medicine. Ultimately, Thompson's critical target is statistical methods *alone* as reliable tools for causal inference. His argument largely hinges upon the crucial differences between trials in biomedical contexts and in agricultural settings, where Fisher first developed the methods of randomization. He thus emphasises the role of theory and of background knowledge in establishing causal claims. Thompson's emphasis on the role of 'non-statistical' elements in causal inference is also shared by Alex Broadbent in *Inferring causation in epidemiology: mechanisms, black boxes, and contrasts* and by Harold Kincaid in *Causal modelling, mechanism, and probability in epidemiology*. They turn attention to the contentious issue of whether causal claims in epidemiology are supported by mechanisms and, if so, how. Broadbent in particular opposes the 'mechanistic

stance' and the 'black box stance' in epidemiology. He thoroughly discusses pros and cons of taking mechanisms as necessary or sufficient to establish causal claims. He also investigates assumptions and consequences of taking mechanistic considerations in causal assessment to be descriptive or normative. Kincaid, on the other hand, focuses on the use of mechanisms, hoping to make observational studies in epidemiology 'more formal' and consequently stronger.

In *The IARC and mechanistic evidence*, Bert Leuridan and Erik Weber focus on yet another aspect of using mechanisms. Their philosophical considerations about causality and mechanisms are more specifically applied to the procedures for evaluating carcinogenicity of agents by the International Agency for Research on Cancer (IARC). They argue for an evidential role of mechanisms. Mechanisms help in excluding confounding, that is when one or more variables interfere and confound the 'real' causal relations. This may lead IARC panels to conclude that an agent is carcinogenic when it is not, and vice versa. A more theoretical contribution is that of Donald Gillies in *The Russo-Williamson thesis and the question of whether smoking causes heart disease*. Gillies specifically addresses the thesis, put forward in Russo and Williamson 2007, that evidence of both difference-making and mechanisms are needed to establish causal claims. Using examples from the studies on smoking and heart disease, Gillies refines the thesis, requiring that mechanisms be 'plausible' rather than 'confirmed' or 'well established'.

The leitmotif of the chapters of this part seems to be that (*pace* EBM partisans) there is more to causation in health contexts than simply statistics. This, as we shall see next, is a thread followed also in the investigations on causality in psychology. Likewise, chapters in the psychology part share concerns about the role and import of difference-making and mechanistic information for disease causation or causal assessment. Another relevant aspect highlighted by this sample of works in the health sciences is that debates on conceptual issues such as mechanisms are not pursued in abstract terms but are meant to positively contribute to the discussions about the 'use' of causality, for instance in IARC procedures.

1.2.2 Psychology

Psychology has a history of paying serious attention to the philosophical literature and of valuing rigorous philosophical clarification of the basic concepts and distinctions of psychology. Philosophy has not always returned the compliment. As a result, many philosophers will be unaware of the explosion of work in psychology on all aspects of causal reasoning. This fascinating work should be of interest not only to philosophy but also to *any* area of science that is wrestling with causality.

Psychologists test empirically how people *do* reason, not directly how people *should* reason. Nevertheless, empirical results are of direct interest to other

fields. Psychologists test 'folk intuitions' on causal reasoning, which is a useful check on whether philosophers' intuitions are systematically different from those of the unphilosophical folk. For the rest of science, primarily concerned with the *normative* aspect of causal reasoning, psychology can find out which weaknesses and fallacies we are susceptible to in our causal reasoning, and in which circumstances we do better – or worse.

The first three chapters in this part give a taster of this growing psychological literature. In *Causal thinking*, David Lagnado brings to bear a body of empirical work to criticize the usual practice in psychology of separating the study of causal learning (learning *about* a causal structure) and causal reasoning (reasoning on the basis of a *known* causal structure). Lagnado argues that studying both aspects together – on the basis of a psychological process account of causal reasoning – will be superior. In *When and how do people reason about unobserved causes?*, Benjamin Rottman, Woo-kyoung Ahn and Christian Luhmann similarly use a body of empirical work to argue that people's reasoning about unobserved causes is more sophisticated than has been recognized. Reasoning about unobserved causes is a big problem for inferring causation from correlation – a concern of almost any science. In particular they examine patterns that people infer in data that deviates from simple correlations and conclude that there is a dynamic interplay between observed and unobserved causes that any attempt to explain causal learning must consider. Clare Walsh and Steven Sloman, in *Counterfactual and generative accounts of causal attribution*, argue that there is evidence that people think about both counterfactuals and mechanisms in forming causal judgements. They go on to examine reasoning about *prevention* or causing an *absence*, noting that there is considerably less consensus on prevention than on positive causation.

The remaining chapters are more philosophical, and illustrate how integrating psychological and philosophical work can benefit both disciplines. In *The autonomy of psychology in the age of neuroscience*, Ken Aizawa and Carl Gillett examine the issue psychologists or neuroscientists face when they discover more than one neurological realizer for what was initially treated as a single psychological phenomenon: do they keep a single psychological phenomenon, with multiple realizers, or do they decide that after all there was more than one psychological phenomenon? Aizawa and Gillett argue, with reference to the discovery of the neural realizers of colour vision, that the higher-level theory plays an essential role in this choice. It is worth noting that Baetu's chapter in Section 1.2.4 examines the same theme with regard to classical genetics and molecular biology. Otto Lappi and Anna-Mari Rusanen, in *Turing machines and causal mechanisms in cognitive science*, argue that explanation using abstract representations in Turing machines illustrates limitations of the account of mechanistic explanation put forward in recent years in the philosophical literature on mechanisms. Finally, in *Real causes*

and ideal manipulations: Pearl's theory of causal inference from the point of view of psychological research methods, Keith Markus sets out a detailed examination of Pearl's account of causal reasoning (see Section 1.2.5), when applied to psychology. Markus discusses ways in which Pearl's formalism should be interpreted and argues that it has certain limitations in the context of psychology.

This part thus develops themes arising in psychology itself and from examination of psychology – but which are also of vital interest to other sciences. If it is true that reasoning to a causal structure and reasoning from a causal structure influence each other, as Lagnado argues, then that is of concern to the many scientists whose work is related to either or both forms of reasoning. The issue of scientific taxonomy, or how a field should chop up its domain, is of wide concern, as is examination of the limitations of highly successful inference methods such as those based on Pearl. Finally, mechanisms and their position in causal judgements, and explanations, are clearly of increasing interest. The use of mechanisms in causal reasoning is now a substantial debate in psychology, that has come from philosophy, and – on the basis of much of this volume – is rapidly becoming a debate of interest right across the sciences.

1.2.3 Social sciences

The social sciences are another area that took time to attract the interest of philosophers of causality and of philosophically-minded scientists. It has perhaps been methodological advancements, especially in quantitative research, that have enabled the social sciences to shake off an inferiority complex with respect to the hard sciences. Arguably, the social sciences still cannot establish the same kind of laws as physics, but the use of more rigorous methods has allowed a much deeper understanding of many social phenomena, and more accurate predictions and well-informed interventions through social policy. Moreover, recent debates in philosophy give room for rethinking even traditional debates in social science.

This is the case, for instance, in the contributions *Causal mechanisms in the social realm* by Daniel Little and *Getting past Hume in the philosophy of social science* by Ruth Groff. On the one hand, Little endorses causal realism and asks what ontology is to be developed for the social realm. He argues for mechanisms, but within a microfoundations approach: in social contexts 'causal mechanisms are constituted by the purposive actions of agents within constraints'. Little also makes clear that such an ontology in the social context is overtly anti-Humean, because causation is not in regularities but in mechanisms. Humeanism is also the critical target of Groff. Notably, she discusses how the tacit Humean metaphysics can be by-passed in social science and touches on issues related to methodological individualism and causal powers. Although Groff does not offer any definite positive arguments, she nicely builds bridges between the traditional philosophy of science literature, stances in analytic philosophy, and the methodology of social science. The

kind of anti-Humeanism argued for in these two chapters concerns metaphysics, namely whether or not all there *is* about causation is the regular sequence of effects following causes in time. The line of argument of Little and Groff may also be extended to epistemological considerations, namely whether or not in order to *know* about causal relations all we have to do is to track regular sequences of effects following causes in time. An attempt to challenge 'epistemological Humeanism' has been carried out by Russo and Williamson (2009a, b) for the social sciences and for epidemiology. Russo has argued that causal epistemology hinges upon the notion of variation. Simply put, model building and model testing is about meaningful joint *variations* between variables of interest; conditions of invariance of parameters or regularity of occurrence are instead constraints to ensure that variations are causal rather than spurious or accidental. Arguments given in Section 1.2.2 seem to suggest that psychologists also track variations rather than regularities.

In the next group of chapters, two main issues come up: mechanism and structure. In *Causal explanation: recursive decompositions and mechanisms*, Michel Mouchart and Federica Russo tackle the problem of causal explanation in social science research, especially quantitative-oriented research. They present the structural modelling approach as a means to causally explain a social phenomenon and advance the view that the core formal tool – i.e. the recursive decomposition – needs to be interpreted in mechanistic terms. In *Counterfactuals and causal structure* by Kevin Hoover, structural modelling has a slightly different facet. 'Structural' does not refer to the structure or *mechanism* that the recursive decomposition represents, but to the structural *equations*. Hoover's structural account hinges on Simon's notion of causal ordering, and the key aspect is the invariant parametrization of the system. The two chapters have in common, though, that structural modelling is an alternative to a manipulationist or interventionist account *à la* Woodward (Woodward, 2003). Simply put, manipulationist accounts hold that x is a cause of y if, and only if, were we to manipulate or intervene on x, some change in y would accordingly follow (with the usual caveats of holding fixed any other factor liable to interfere in the relation between x and y). Interestingly enough, manipulability theories now enter philosophical discussions in a different way. It seems that the importance of the notion of manipulation is not so much in providing an explication of the concept of causation, but rather in explicating other notions, e.g. that of 'constitutive relevance' used by Craver (2007). In *The error term and its interpretation in structural models in econometrics*, Damien Fennell also considers structural models based on Herbert Simon's notion of causal ordering, and in particular examines issues related to the error term in the equations. The goal of the chapter is mainly expository, in making those who use these kinds of models in econometrics aware of conceptual issues that can hinder successful and meaningful results. In the last chapter *A comprehensive causality test based on the singular spectrum analysis*,

Hossein Hassani, Anatoly Zhigljavsky, Kerry Patterson and Abdol S. Soofi discuss a new statistical method for testing causal relations not in the tradition of structural modelling, but rather in the tradition of Granger's approach. In this approach causality does not lie in the *structures* or *mechanisms* identified in the joint probability distributions, but, to put things very simply, it lies in the power of a (set of) cause-variables to convey information in order to predict the effect-variable.

The second group of chapters, and in particular Hoover's, is closely related to issues also addressed by Judea Pearl and Nancy Cartwright in Section 1.2.5: counterfactuals and structural models, structural models and external validity, going beyond statistics in drawing causal inferences. It is also worth mentioning that most chapters again deal, either directly or indirectly, with mechanisms. A possible explanation is that one may require more than probabilities to give a satisfying conceptual analysis of causation. Perhaps probabilities are not enough even from a methodological point of view: large parts of these methodological chapters invoke, albeit in different ways, mechanisms. The fact that so much emphasis is given to mechanisms may be due to a shift of focus from probabilities to mechanisms. This does not necessarily mean, of course, that probabilities do not play any role in the explication of causation.

1.2.4 Natural sciences

The natural sciences, and particularly physics, are the traditional source for philosophers of science, and for many years natural scientists have taken an interest in the philosophical literature on their field. This part begins with chapters representing the established but growing interest of philosophers in the biological sciences. These engage with topics from mechanism discovery in molecular biology to mathematical modelling in evolutionary biology. The increasing diversity of engagement between philosophy and the natural sciences is also represented by work on the far newer climate science. The part closes with two chapters demonstrating the cutting edge of work on causality emanating from physics.

In *Mechanism schemas and the relationship between biological theories*, Tudor Baetu looks at the relationship between classical genetics and molecular biology, and argues that there are cases where the accommodation of data from molecular biology results in better *classical* explanations. For example, Marfan, Loeys-Dietz and Ehlers-Danlos syndromes can be confused as a single genetic disease, but they are different – and this has implications for their treatment. Baetu argues that this means that classical genetics and molecular biology are not merely parallel explanatory projects, but related. He offers an account of this relation in terms of mechanism schemas. Note that this chapter is thematically linked to the chapter by Aizawa and Gillett in Section 1.2.2. The common concern is with when a difference in lower-level realiser (for Baetu, biochemical molecules; for Aizawa and Gilett,

neural systems) matters to the higher-level theory (for Baetu, classical genetics; for Aizawa and Gillett, colour vision), when it does not, and why.

Roberta Millstein turns to our concept of *chance* in *Chances and causes in evolutionary biology: how many chances become one chance*. Millstein argues that at least seven colloquial uses of chance in evolutionary biology can all be translated into the Unified Chance Concept (UCC) by specifying the types of causes that are taken into account (i.e. considered), the types of causes that are ignored or prohibited, and the possible types of outcomes. The UCC is useful, Millstein argues, because it makes it easier to translate between the colloquial chance concepts, and also from them to more formal probabilistic language. In *Drift and the causes of evolution*, Sahotra Sarkar takes a very different approach. Drift is an explanation for evolutionary outcomes which are *not* due to natural selection, mutation, migration or the other recognized causes of evolution. There is always deviation from expected outcomes due to these causes, and this is drift. Sarkar works in the framework of mathematical modelling of evolutionary processes. He distinguishes between the constitutive and the facultative assumptions of a model. The constitutive assumptions define the model, and cannot be changed without changing the system, while the facultative assumptions can vary. So the facultative assumptions give you the causes which act against the background conditions that are given in the constitutive assumptions. Sarkar takes whether the initial size of a population is finite or infinite as a constitutive assumption, and builds a simple mathematical model to show that this models drift, satisfying the usual conditions for drift. But drift is in no facultative assumption of this model. All that is required for drift is that the population be of a finite size; this finite size is part of the conditions under which the evolutionary causes – selection and mutation – operate. Sarkar concludes that drift is not a cause of evolution.

In the chapter, *In defense of a causal requirement on explanation*, Garrett Pendergraft examines whether equilibrium explanations, which explain an observed equilibrium state of a dynamical system by providing a range of possible initial states and possible causal trajectories of the event being explained, violate Pendergraft's Causal Factors Requirement: an explanation of an event must provide information about the causal factors that influenced whether or not that event occurred. Pendergraft argues that equilibrium explanations satisfy this, since they do provide information about causal factors. In so far as drift is an explanation of evolutionary outcomes in terms of chance, the question of whether or not it is a cause of evolution is a link between Millstein and Sarkar's work, and Pendergraft's.

Paolo Vineis, Aneire Khan and Flavio D'Abramo, in *Epistemological issues raised by research on climate change*, examine some of the epistemological challenges faced by climate change research. This is an area with special challenges for coming to causal conclusions, since randomized experiments cannot be done, but only experiments on microenvironments artificially

constructed in the laboratory, where the results don't always extrapolate to the real world, and some highly speculative attempts to control real weather, such as to make rain by seeding clouds. This chapter considers particularly the example of rising levels of certain diseases that can clearly be traced to rising salt levels in Bangladesh, and whether we can say that climate change caused these diseases.

One interesting account of causality that emanates from the traditional engagement of philosophy with physics is the process theory of causality (Reichenbach 1956; Salmon 1998; Dowe 2000). Reichenbach's seminal idea, taken up and developed by Salmon, held a process to be causal if it is capable of transmitting a mark. Salmon and Dowe later adopted a version of the theory according to which a causal process is one that transmits or possesses a conserved physical quantity such as charge or angular momentum. In *Explicating the notion of 'causation': the role of extensive quantities*, Giovanni Boniolo, Rosella Faraldo and Antonio Saggion present a development of the process theory, in which conserved quantities are replaced by *extensive quantities*. An extensive quantity is defined as a quantity whose value is given by the volume integral of some function defined over space-time points. Extensive quantities include conserved quantities like angular momentum and charge, but also quantities such as volume and entropy.

For Reichenbach, causal relationships were also characterized *probabilistically*. His probabilistic theory was based around the *common cause principle*, which says roughly that if two events are probabilistically dependent but neither causes the other, then there is some set of common causes of the two events that screens off the dependence (i.e. the two events are probabilistically independent conditional on the common causes). Miklós Rédei and Balázs Gyenis, in *Causal completeness of probability theories – results and open problems*, investigate the question of when the common cause principle is satisfiable. It turns out that in some probability spaces it is possible to satisfy the principle but in others it is not. Their chapter considers both classical and non-classical probability spaces and presents the state-of-the-art concerning what is known about this problem.

On the surface these chapters are very different, arising from different concerns from different scientific fields. But there are some common themes at work here, and in the rest of the volume. The concern of Vineis, Khan and D'Abramo over difficulties with randomized experiments also arises in Section 1.2.1, on the health sciences, and in Section 1.2.5, on computer science, probability and statistics. The issue of mechanisms arises here, as elsewhere. For Baetu, understanding mechanisms and mechanism discovery is vital to understanding the relation between theories, while for Pendergraft the challenge is better to understand different approaches to explanation. The overall project of better understanding explanation is also reflected in the chapter by Lappi and Rusanen examining mechanistic explanation in Section 1.2.2.

The work on mechanisms is, on the face of it, very different from the work on causal processes originating in physics, but there are commonalities in the role of mechanisms and processes in causal explanation and inference, as developed in Section 1.2.6.

1.2.5 Computer science, probability and statistics

As discussed earlier, in the face of criticisms from Mach, Pearson and Russell, in the twentieth century research scientists largely avoided explicit discussion of the causal claims that were implicit in their papers. But certain developments at the turn of the millennium have helped to rehabilitate explicit talk of causality in the sciences, and now 'causality' is no longer a dirty word. It is in the context of these developments that the chapters of this part of the book should be placed.

The 1980s saw the beginning of a revolution in the use of causal methods in the sciences, stemming from interest amongst computer scientists and statisticians in probabilistic and graphical methods for reasoning with causal relationships. Of course revolutions don't just pop out of thin air, and there were several – rather disjoint – lines of thought that led to these important advances. Notably, philosophers of science attempted to characterize causal relationships in terms of patterns of probabilistic dependencies and independencies, and represent them graphically using 'causal nets' (Reichenbach 1956); computer scientists used graphs that chart probabilistic dependencies and independencies to construct computationally tractable representations of probability distributions (see, e.g. Chow and Liu 1968); statisticians were also using graphical models to represent dependence and independence relationships in the analysis of contingency tables (Darroch *et al.* 1980). In the 1980s these advances led to *Bayesian net* methods for causal reasoning (Pearl 1988). Here causal relationships are represented by a directed acyclic graph and causality is tied to probability via the *causal Markov condition*, which says that each variable in the network is probabilistically independent of its non-effects, conditional on its direct causes (see, e.g. Williamson 2005). In the 1990s these methods were reconciled with the use of structural equation models to handle causal relationships – a formalism, stemming from work in the 1920s, that is essentially very similar to the Bayesian net approach (Pearl 2000). As can be seen from the chapters in this part of the book, the Bayesian net approach, and more generally the approach to causality stemming from recent developments in computer science, probability and statistics, remains a thriving area of interesting research questions and lively debate.

In *Causality workbench*, Isabelle Guyon, Constantin Aliferis, Gregory Cooper, André Elisseeff, Jean-Philippe Pellet, Peter Spirtes and Alexander Statnikov focus their attention on methods for the automated learning of causal models directly from data. Hitherto, the field of machine learning in computer science has primarily concerned itself with the task of generating

models that are predictively accurate. Broadly speaking, predictive accuracy merely requires that the model adequately capture the underlying probability distribution. Recently, however, there has been some demand for models that are explicitly causal, in order to predict the effects of interventions. Thus a supermarket may wish to use shopping data not only to predict which aisles will need stocking most regularly but also to determine where to move a particular product in order to increase sales of that product. *Causality workbench* presents and discusses an exciting new testbed for computer systems that attempt to learn causal relations directly from data.

The standard approach to learning causal relationships from data is to find a Bayesian net with the least number of arrows from all those that fit the data, and to interpret the arrows in the graph of that net as characterizing the causal relationships. In *When are graphical causal models not good models?*, Jan Lemeire, Kris Steenhaut and Abdellah Touhafi argue that this approach may be unsatisfactory. By appealing to ideas concerning *Kolmogorov complexity*, used widely in computer science in the context of data compression, they argue that the correct causal model may not be a minimal Bayesian net.

Under the standard machine learning approach, the probabilities of a Bayesian net that has been learnt from data are usually simply the frequencies induced by the data. But Bayesian nets were originally conceived of as *belief networks*: the probabilities in the net were supposed to represent degrees of belief that would be appropriate for an agent to adopt given the evidence of the data (Pearl 1988). In *Why making Bayesian networks objectively Bayesian makes sense*, Dawn Holmes argues for a return to the Bayesian, degree of belief interpretation. But rather than advocating the usual subjective Bayesian approach, according to which degrees of belief are subject to rather loose constraints and are largely a question of personal choice, Holmes advocates objective Bayesianism, which holds that degrees of belief are typically subject to tight constraints that leave little or no room for personal choice (Jaynes 1957; Williamson 2010). The key question is: given certain causal and probabilistic evidence, which Bayesian net best represents appropriate degrees of belief? This question has been tackled by Williamson (2005) and Schramm and Fronhöfer (2005), as well as in a distinct line of work culminating in Holmes' chapter.

Bayesian nets are normally construed as representing causal relationships in a qualitative way, via the arrows in the graph of the net. But one might suspect that causality is a matter of degree, in which case the question arises as to how one could measure the extent to which one variable causes another. This question is taken up by the next two chapters. *Probabilistic measures of causal strength*, by Branden Fitelson and Christopher Hitchcock, presents a detailed comparative analysis of a plethora of measures of causal strength that have been put forward in the literature on causality. Kevin Korb, Erik Nyberg and Lucas Hope, in their chapter, *A new causal power theory*, argue that

a good measure of degree of causal power can be constructed by appealing to concepts from information theory in computer science – in particular to the concept of mutual information, a concept that is very natural in this context and which underpins, for example, the approach of Chow and Liu (1968) alluded to above.

A quantitative view of causal relationships also forms the backdrop of *Multiple testing of causal hypotheses* by Samantha Kleinberg and Bud Mishra. Their chapter seeks to use methods from computer science and statistics to determine those causal hypotheses that are significant in the statistical sense. Rather than framing their approach in the Bayesian net formalism, which can struggle to cope with the kind of time-series data under consideration in this chapter, Kleinberg and Mishra develop a framework using other methods from computer science, in particular temporal logic and model checking. They apply their approach to microarray data, to neural spike trains, and also to data concerning political speeches and job approval ratings.

Machine learning methods for constructing Bayesian nets can be categorized according to whether or not they attempt to discover latent variables, i.e. variables which are not themselves measured in the data but which are causes of two or more variables that are measured. Latent variables are important to many sciences, not least to psychology which typically uses factor analysis to discover unmeasured common causes (cf. the chapter of Rottman *et al*. discussed in Section 1.2.2). In *Measuring latent causal structure*, Riccardo Silva presents an approach to learning causal relationships that explicitly represents latent variables as nodes in the graph of the Bayesian net. This approach is applied to an example concerning democracy and industrialization and to an example concerning depression.

Judea Pearl, in *The structural theory of causation*, continues his programme of providing a mathematical formalism for causality that unifies approaches to causal reasoning that are extant in the sciences. After explaining the core features of his new theory – which extends the Bayesian net approach of Pearl (1988) and the structural equation approach of Pearl (2000) – Pearl discusses how his new theory can underwrite counterfactual conditionals (conditionals whose antecedents are false), a topic already encountered in Hoover's chapter earlier in the volume. Pearl argues that his account supersedes attempts by philosophers of science to provide a probabilistic analysis of causality, and should be preferred to the *potential-outcomes* (also called *potential response*) approach that emerges from work by Neyman and Rubin (Neyman, 1923; Rubin, 1974).

The potential response approach is also discussed by Sara Geneletti and Philip Dawid in *Defining and identifying the effect of treatment on the treated*. They argue that their decision-theoretic version of the Bayesian net approach can be viewed as a generalization of the potential response approach. Moreover, they argue that their approach can be used to formulate and measure the

effect of treatment on the treated, an important measure of causal strength that applies to cases where those who are treated are to some extent self-selected.

Statistical and machine learning methods examine data involving a sample of individuals and make general causal claims on the basis of this data. (As Geneletti and Dawid emphasize, one needs to be very careful not to overgeneralize at this stage.) Then policy makers need to apply the general causal claims to a group of individuals who require remedial action in order to identify the most effective interventions. This two-stage process is the focus of Nancy Cartwright's chapter, *Predicting 'it will work for us': (Way) beyond statistics*. Cartwright argues that statistical methods alone will not guarantee the success of either stage of the process. The second stage, Cartwright maintains, needs to be informed by case-specific causal models, concerning the group of individuals who will be treated, and this requires local knowledge about that group that goes well beyond the original dataset. On Cartwright's account, general causal claims are claims about tendencies or capacities and the first stage needs to be backed up by theoretical knowledge of the domain – knowledge of the mechanisms that are responsible for the regularities in the data. (This latter view accords with Thompson's chapter, discussed in Section 1.2.1.)

1.2.6 Causality and mechanisms

This final part of the volume examines mechanisms and their relationship to causality. Mechanisms are important to causal *explanation*, as one way of explaining a phenomenon is to point out the mechanism responsible for it. As we see in the parts on Health Sciences, Section 1.2.1, and Social Sciences, Section 1.2.3, mechanisms are of increasing importance in causal *inference*. (See also Russo and Williamson, 2007; Russo and Williamson, 2011; Illari, 2011). As we see in the part on Psychology, Section 1.2.2, mechanisms are also important in causal *reasoning* (reasoning from a known causal structure). It seems that mechanisms are of interest to every aspect of thinking about causality. The widespread feeling that investigating mechanisms is a fruitful avenue to explore is illustrated in the sheer number of chapters in this volume that touch on the methodology, epistemology and metaphysics of mechanisms in some way or another.

This is a very clear case where philosophical theorizing about the methodology of science is of interest right across scientific disciplines. And while these chapters are more theoretical than those in other sections, since they are not examining an issue arising in a single scientific field, they are all aiming to contribute to scientific work, and are scientifically informed. These chapters illustrate the sheer breadth of interesting work concerning causality and mechanisms, stretching from the very idea of mechanism, their metaphysics, and the applicability of particular conceptions of mechanism across scientific domains.

In *The idea of mechanism* Stathis Psillos disentangles two historical ideas of mechanism. The first is the mechanical conception of mechanism, that mechanisms are configurations of matter in motion subject to mechanical laws. Psillos examines Poincaré's critique that such mechanical mechanisms are too easy to envisage to be informative, because if there is any possible configuration of matter in motion that can underpin a set of phenomena, then there is an infinity of such configurations. The second idea of mechanism is the quasi-mechanical conception of mechanisms, where a mechanism is any arrangement of parts into wholes in such a way that the behaviour of the whole depends on the properties of the parts and their mutual interactions, where this is what constitutes their unity. Psillos discusses Hegel's critique that the unity that such mechanisms possess is external to them, because of the need to identify a *privileged* decomposition out of those available, and so the idea that all explanation is mechanical in this sense is devoid of content. Psillos argues on the basis of these two critiques that mechanisms are not the building blocks of nature, so undermining the *metaphysics* of mechanisms, but that nevertheless the search for mechanism is *epistemologically* and *methodologically* useful. This is a valuable critical historical introduction to the idea of mechanism, against which many of the other chapters can be seen as developing a distinct *new* notion of mechanism.

The first two chapters engage with the metaphysics of mechanisms. In *Singular and general causal relations: a mechanist perspective*, Stuart Glennan examines the relation between singular and general causal relations – the difference between Fred's taking penicillin curing *him*, and penicillin in general curing certain forms of infection. Glennan argues that the simplest reason for preferring singularism from a *mechanista's* perspective (the perspective of someone promoting mechanisms for at least one of the three purposes outlined above) is because mechanisms are particulars – particular things. Glennan then argues for singular determination, which is the view that any causal interaction is a singular case of causal determination, where any causal generalisation is true merely in virtue of a pattern of such singular instances. For Glennan, this is the best metaphysical view of the fundamental *components* of mechanisms since it offers a unified singularist view of these, with the singularist view of mechanisms *themselves* as particulars. Phyllis McKay Illari and Jon Williamson, in *Mechanisms are real and local*, examine the implications of two widely shared premises concerning mechanistic explanation: that mechanistic explanation offers a welcome alternative to traditional laws-based explanation, and that there are two senses of mechanistic explanation: epistemic and physical explanation. They argue that in mechanistic explanation, mechanisms are treated as both real and local, and argue that reality and locality require an *active* metaphysics for the components of mechanisms, illustrated using Cartwright's capacities approach.

The next two chapters both address the idea of a causal *link*, or causal continuity. In *Mechanistic information and causal continuity*, Jim Bogen and Peter Machamer set out to give a novel account of causal continuity in terms of mechanistic information. They use examples of Crick's early conception of gene expression and a sensory-motor reflex in the leech to argue that mechanistic information can be understood in terms of *goals* served by mechanisms, and the *reach* or strength and independence of influence of initial stages of the mechanism on the final stages. Information is ineliminable because the continuity of some mechanisms is a function of their teleological structure, i.e. the goal of the mechanism, and so without attention to the teleological structure, the vital continuity is lost. This chapter has potential implications for the epistemology and methodology of mechanisms, along with their metaphysics. In *The causal-process-model theory of mechanisms*, Phil Dowe addresses the issue of the applicability of causal process theories – such as his own view, mentioned in Section 1.2.4, that causal processes involve the maintenance of conserved quantities – to areas of science other than physics. Dowe considers the need for an account of what it is that scientists look for when they look for something that underlies correlations as an important motivation for his account. If processes involve a spatiotemporally continuous link between cause and effect, then processes cannot involve absences, which would be a gap in a causal process. But absences are sometimes cited as cause or effect, such as in: 'my failure to water the plants caused their death'. An *absence* of watering is said to cause a positive outcome. Dowe offers an account of causal relevance in mechanisms, which can incorporate his theory that causation involves causal processes understood in terms of conserved quantities, but which also allows *absences* in causal explanation.

Meinard Kuhlmann, in *Mechanisms in dynamically complex systems*, examines whether the concept of mechanism can be extended to cover systems that are not just *compositionally* complex, but exhibit complex *dynamics* – what he calls 'dynamically complex systems'. These dynamics arise from the interaction of the system's parts, but are largely irrespective of many properties of these parts. Kuhlmann uses detailed examples of dynamical systems in analysis of heart beat, and financial markets, to argue that dynamically complex systems are not sufficiently covered by the available conceptions of mechanisms. He explores how the notion of a mechanism has to be modified to accommodate this case.

Julian Reiss, in *Third time's a charm: causation, science, and Wittgensteinian pluralism*, examines pluralism about causality: the claim that there is no single correct account of what *cause* means, but instead multiple concepts of cause. Reiss examines three different accounts that all reject any attempt to define 'cause' in terms of necessary and sufficient conditions. Instead they regard different instances of causal relationships such as 'pulling', 'pushing', 'breaking' or 'binding' as sharing family resemblances at best: pushing and

pulling clearly share something in common, as do breaking and binding, but there is no single property shared by all instances of such causal terms. This is a pluralist tradition inspired by Wittgenstein and shared by Anscombe, Cartwright, and Machamer, Darden and Craver, and is a form of pluralism about causality that interests many working on mechanisms. Reiss argues for the third form of pluralism, which he says is a form of inferentialism: the method of verifying a causal claim – of evidentially supporting it – determines with what other claims it is inferentially related.

In different ways these chapters are attempting to give an account of mechanisms suitable to their place in causal explanation, inference and reasoning. It is their place in explanation that drives Glennan's emphasis on singularism, and Illari and Williamson's related examination of locality, while Machamer and Bogen, and Dowe's very different attempts to give an account of the causal link – if successful – are important to the usefulness of mechanisms to causal inference. One ambition is also for a single account of mechanism that is applicable across scientific disciplines. Ultimately, the hope is for a general account of mechanisms – the first glimmerings of which can be seen here – which fruitfully addresses all three *methodological* uses of mechanisms in all scientific *disciplines*. This is ambitious, and it remains an open question whether it will be possible.

1.3 Whence and whither?

Progeny of the chapters.
Some of the chapters in this volume were invited contributions, but most were submissions to an open call for papers. Within the broad remit of *Causality in the Sciences*, all authors chose their own topics and titles, and all papers were refereed. Many submissions were received from participants in two events of the *Causality in the Sciences Conference Series* (`http://www.kent.ac.uk/reasoning/cits`): *Causality Study Fortnight* held at the University of Kent in September 2008, and *Mechanisms and Causality in the Sciences* held at the University of Kent in September 2009.

Next steps.
The individual chapters in this volume indicate a plethora of open questions for research on causality. Here we highlight just a few topics for future research that stand out as particularly pressing.

From the volume it is clear that there is a mature field of research centred on the question of the relationship between causality and probability (see Section 1.2.5). But the volume also indicates that there is also a newer, rapidly developing area of research, exploring the relationship between causality and mechanisms (see in particular Section 1.2.6). However, we received very few

papers on *all* three: on causality *and* probability *and* mechanisms, and the question of how probabilistic accounts of causality can mesh with mechanistic accounts of causality desperately needs answering. This suggests that a first hot topic for future research will be on causality, probability and mechanisms, bridging the causality-probability agenda on the one hand, and the causality-mechanisms agenda on the other hand.

Although successful formalisms exist for handling aspects of causal reasoning using probabilities, few are explicitly designed for handling mechanisms (see however the discussions of the possible mechanistic interpretations of models in social research in Section 1.2.3). Indeed, a detailed formal understanding of causal reasoning using mechanisms is sorely lacking. So a second hot topic is likely to concern formalisms for handling mechanisms, particularly in causal inference and reasoning. Such formalisms may emerge from the existing formalisms for reasoning using probabilities (e.g. Bayesian nets or multilevel models), or they may need to be entirely new – tailor-made methods for handling causal mechanisms.

A number of chapters invoke mechanisms as evidence for causal tasks, e.g. for the assessment of carcinogenicity. Interestingly, in biomedical and social contexts alike scientists are suggesting that the 'mechanistic picture' is more complicated than it may look at first sight. They are thus moving towards 'ecological views', namely approaches that aim to include both biological and socio-economic factors in the *same* mechanism. This suggests a third hot topic will be to develop *pan-scientific* causal methods. In particular, we are in need of accounts where (i) the *concept* of mechanism permits the inclusion of factors of different natures, (ii) factors of different natures can provide multiple points of *epistemic access* to the same mechanism, and (iii) *formal models* can handle factors of different natures.

Having presented three questions that are likely to feature in future research, we should make some cautionary remarks about how these questions might be solved. We suggested in our manifesto (Section 1.1) that theorizing about causality is best pursued as a collaborative project involving both philosophers of science and scientists from different disciplines and fields. But such a broad project poses two related challenges.

Causality is at the crux of metaphysical, epistemological and methodological issues in the sciences. And different participants in the debate have different primary concerns. The first challenge in theorising about causality is to avoid blurring these three kinds of issue, remaining explicit about which kind is being addressed, and how. For example, the question above of how to integrate ontologically different factors in the same mechanism has metaphysical, epistemological and methodological facets. Yet giving a *methodological* answer to someone concerned about the *metaphysics* of this question, or vice versa, will not help them.

Nevertheless, the metaphysics, the epistemology and the methodology of causality are not wholly distinct. We should expect answers to any one of the three kinds of issue to have implications for the other two kinds. The second challenge is to produce an understanding of causality that successfully addresses all three kinds of issue in a unified way, *without* blurring the distinctions between metaphysics, epistemology and methodology. To make progress on this requires making explicit how metaphysics, epistemology and methodology impact on each other. This is challenging. Note that Cartwright (2007) is pioneering in this regard, urging that questions of metaphysics, methods and *use* cannot be successfully addressed in isolation. Cartwright makes it clear that she thinks an *understanding* of causality that does not help us address how causal claims inform *policy* will never be adequate.

In an era of concern about the 'impact' of research, philosophers have to make the effort to explain why and how philosophical discussions of causality have a bearing on policy and other questions of intervention and control. But scientists also need to make an effort to step back and think of the coherence of the foundations of their work: a 'methodological salad' – an eclectic mix of methods – will inspire no confidence at all unless unifying foundations can be found for the ingredient methods.

In sum, while a sound understanding of causality can best be gained though a mutual project involving the sciences and philosophy, care must be taken not to make progress on metaphysics at the expense of epistemology and methodology, or vice versa.

Acknowledgements

This volume is a product of two research projects hosted by the Centre for Reasoning at the University of Kent: *Mechanisms and causality*, funded by the Leverhulme Trust, and *Causality across the levels: biomedical mechanisms and public health policies*, funded by the British Academy. We are also grateful to the funders who supported the *Causality in the Sciences* conferences that gave rise to some of the papers of this volume: the British Academy, the Mind Association, the British Society for the Philosophy of Science, the Aristotelian Society, the Kent Institute for Advanced Study in the Humanities, the School of European Culture and Literature and the Centre for Reasoning at the University of Kent. We also wish to thank the other members of the steering committee of the *Causality in the Sciences* conference series: Julian Reiss and Erik Weber, and all the participants at those conferences for the lively debates. Finally, we are deeply indebted to all the authors and referees for their hard work on this volume, and to Elizabeth Hannon, Dewi Jackson, Louise Kane, Keith Mansfield, Victoria Mortimer and Mike Nugent at Oxford University Press.

References

Cartwright, N. (2007). *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press, Cambridge.

Chow, C. K. and Liu, C. N. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, **IT-14**, 462–467.

Craver, Carl (2007). *Explaining the Brain*. Clarendon Press, Oxford.

Darroch, J. N., Lauritzen, S. L., and Speed, T. P. (1980). Markov-fields and log-linear models for contingency tables. *Annals of Statistics*, **8**, 522—539.

Dowe, Phil (2000). *Physical Causation*. Cambridge University Press, Cambridge.

Illari, Phyllis McKay (2011). Mechanistic evidence: Disambiguating the Russo-Williamson thesis.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, **106(4)**, forthcoming.

Neyman, Jerzy (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, **5**(4), 465–472. 1990. Translated by Dorota M. Dabrowska and Terence P. Speed.

Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo CA.

Pearl, Judea (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

Reichenbach, Hans (1956). *The Direction of Time* (1971 edn). University of California Press, Berkeley and Los Angeles.

Rubin, Donald B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66**(5), 688–701.

Russo, Federica (2009*a*). *Causality and Causal Modelling in the Social Sciences. Measuring Variations*. Methodos Series. Springer, New York.

Russo, Federica (2009*b*). Variational causal claims in epidemiology. *Perspectives in Biology and Medicine*, **52**(4), 540–554.

Russo, Federica and Williamson, Jon (2007). Interpreting causality in the health sciences. *International Studies in Philosophy of Science*, **21**(2), 157–170.

Russo, Federica and Williamson, Jon (2011). Generic versus single-case causality: the case of autopsy. *European Journal for Philosophy of Science*, **forthcoming**.

Salmon, Wesley C. (1998). *Causality and Explanation*. Oxford University Press, Oxford.

Schramm, Manfred and Fronhöfer, Bertram (2005). Completing incomplete Bayesian networks. In *Proceedings of the Workshop on Conditionals, Information and Inference* (ed. G. Kern-Isberner, W. Rödder, and F. Kulmann), Lecture Notes in Artificial Intelligence 3301, pp. 200–218. Springer, Berlin.

Williamson, Jon (2005). *Bayesian Nets and Causality: philosophical and Computational Foundations*. Oxford University Press, Oxford.

Williamson, Jon (2010). *In Defence of Objective Bayesianism*. Oxford University Press, Oxford.

Woodward, J. (2003). *Making Things Happen: a Theory of Causal Explanation*. Oxford University Press, Oxford.

PART II
Health sciences

This page intentionally left blank

·2·
Causality, theories and medicine

R. Paul Thompson

Abstract

Randomized controlled trials (RCTs) are pervasive in clinical medical research, which stands in stark contrast to other sciences such as physics, chemistry and biology. Most clinical researchers that use RCTs regard them as uncovering causal connections. R. A. Fisher best articulated the rationale for this position in 1935. According to Fisher, if randomization, blocking and replication demonstrated a connection between an intervention and an outcome, that connection is causal. In this paper, I argue that RCTs in clinical medicine do not reveal causal connections. Causal claims in clinical medicine, as in the rest of science, are justified by reference to a robust theory, not RCTs. Part of the argument rests on crucial differences between Fisher's use of RCTs in agriculture and the current use of RCTs in clinical medicine. Two key differences are: the different role of randomization and the legitimacy of assuming homogeneity of the intervention and control entities. A more significant part rests on the integrative power of robust theories; causal attributions are justified by demonstrating that they are, or can be, embedded in a large well-confirm framework. RCTs, by contrast, at best provide isolated input-output connections. A secondary thesis of the paper is that robust theories also allow causal claims to be well-confirmed.

2.1 Introduction

R.A. Fisher was a brilliant mathematician whose contributions to the mathematical foundations of statistics were deep, elegant and robust. His applications of statistics to agricultural research and to Mendelian-based population dynamics were, and remain, transformative. In medicine, the nearly ubiquitous acceptance of the mantra that Randomised Controlled Trials (RCTs) are the gold standard of evidence is traceable to Fisher – although most who espouse the mantra seem not to know its Fisherian origins. Embedded in Fisher's conceptions of experimental design and statistical inference is a conception of causality, and of the role and power of randomization. Both these conceptions I shall argue are untenable, especially in the context of medicine.

RCTs may have considerable utility in a number of research contexts since they do provide some support for a scientific theory (a dynamical

(and mechanistic) system which is asserted to model (mathematically) the ontology[1] and dynamics of phenomena). They, however, are far from a gold standard. Contrary to Fisher, they do not provide a causal account of the phenomena under study; nor is randomization essential or, in some contexts, desirable. In what follows, I shall explore what kind of empirical support RCTs provide and why in medicine that support is problematic. I shall also argue that causal attributions are only possible when a robust scientific theory underwrites the attributions; RCTs fall far short of providing a basis for causal attributions. Notwithstanding the ubiquitous use of RCTs in clinical medicine, impressive and robust theoretical underpinnings exist for a substantial array of medical knowledge and causal attributions; it is on these, and not RCTs, that explanation and prediction rest.

2.2 Causality and randomized controlled trials

Outcomes from RCTs constitute the basis for many knowledge claims in medicine and many clinical decisions – especially pharmacological interventions and lifestyle interventions (changes in diet, exercise, and so on). During the last half-century, RCTs have risen in number and authority; government regulators, many epidemiologists and the media have come to regard RCTs as the gold standard of evidence in clinical medicine. This rise in number and authority began with R. A. Fisher who claimed that RCTs provide the basis for discovering and justifying causal connections.[2] According to Fisher, randomization, replication and control (for him, the method of pairwise blocking) guarantee that the intervention is the **cause** of the difference between the outcome in the experimental group (those receiving an intervention) and that in the control group.

Fisher developed his views on experimental design while working at Rothamsted Experimental Station, which he joined as a statistician in 1919. The Station, by that time, had been engaged in agricultural research for more than 70 years. By 1935 when his *Design of Experiments* was published his adherence to the principles of randomization, replication and blocking were

[1] The ontology of a system is the collection of entities postulated to exist (e.g. DNA, proteins, amino acids, leukocytes, and the like) and their properties and their physical relationships to each other (e.g. proteins are strings of amino acids). Dynamical relationships are expressed by a set of equations with these entities as variables.

[2] Randomized trials existed prior to Fisher. James Lind, for example, conducted controlled trials in the eighteenth century – not quite randomized but employing the same reasoning. The most well known is his controlled trial involving the causes of scurvy – a small sample size but manifesting the principles of current RCTs (see: Lind, 1753). Nonetheless, Fisher looms large in the history of RCTs and he contributed significantly to dogma that RCTs provide a strong basis for asserting causal connections and that they should be the basis for experimental design in agriculture and other fields.

firmly entrenched in his method of experimental design. These principles proved powerful in the context of plant agriculture where replication involved sectioning a field into plots, blocking was pairing adjacent plots that were assumed to be identical in all relevant respects (nutrient content and soil composition for example) and randomization was the designation of one of each pair as the experimental plot by a random process such as flipping a coin. Since there were numerous paired plots, the experimental intervention was replicated many times within a single experiment. If a consistent similar difference was found between each experimental plot and its paired control, constant conjunction of intervention and outcome could be assumed. Since paired plots were assumed to be identical in all relevant respects except for the intervention, the outcome could be declared as *caused* by the intervention. Randomization bolstered the assumption of plot-pair identity by removing any systematic bias in favour of one plot of a pair – such as, always choosing the left plot as the experimental plot or favouring lower elevation.

This kind of agricultural research presents an ideal state of affairs for applying Fisher's method of experimental design. The values of the relevant variables in a modest-sized field, as is most frequently used in research, are reasonably homogeneous (soil Ph, organic-to-inorganic material ratio, clay content, and so on). Moreover, dividing the field into small plots and pairing adjacent plots reduces what little heterogeneity might be found in non-adjacent parts of the entire field. Genetic diversity and trait diversity can be managed in agricultural plants to ensure minimal heterogeneity. The adjacent plots remain adjacent throughout the trial and external factors (such as rainfall, hours of sunlight) will be virtually identical for both plots. In short, homogeneity of relevant factors exists naturally or can be easily produced.

In clinical medicine, things are very different. First, a population of individuals is an extremely heterogeneous collection: a wide variety of genotypes, different environmental histories, different physiological dynamics, different interpersonal contacts, and the like. Second, unlike two adjacent plots of land, pairs of individuals or paired groups of individuals do not remain together during a trial; hence, each individual is exposed to different environmental factors. Consider, for example, the simple difference of diet and timing of meals, which may or may not be relevant and often we do not really know their relevance. In the world of RCTs, the Fisherian principles of experimental design are supposed to tame this heterogeneity. The paired groups of individuals are, ideally, random samples from a population; randomization, in principle, results in groups with identical heterogeneity; for each individual in one group, there will be an individual in the other group with the same characteristics. Hence, it is assumed that although any two individuals chosen randomly can be expected to differ in relevant respects, when a large number of individuals are assigned randomly to two different groups, the groups, taken as collectives, will be homogeneous.

Most texts on statistical methods define 'random sample' as: 'A sample of *n* individuals from the population [the whole set or collection of items about which we want information] chosen in such a way that all possible sets of *n* individuals are equally likely to occur' (Wetherill, 1967). As an assumption underpinning aspects of experimental design and statistical analysis, this definition is essential and potent. In the messy world of RCTs, its potency evaporates. First, even the most careful actual random sampling cannot be known to satisfy the requirement of the definition. Second, for financial, ethical and/or trial management reasons, most 'random' sampling in clinical experiments are not from 'the whole set or collection of items about which we want information' (the population) but an already much reduced collection – those, for example, willing and available to participate, those located near the research centre, those who have a relevant disorder, those who do not have signs or symptoms that suggest the experiment might put them at risk, and the list goes on.

Third, even if two groups satisfied the statistical definition of random samples, internal heterogeneity undercuts the assumption that the groups are identical in relevant respects. Most traits of individuals are quantitative traits (traits that differ in the amount or degree) such as blood pressure and lung capacity. These traits are usually the product of many genes and, importantly, are somewhat environmentally sensitive (for example, the effect of exercise on lung capacity and muscle strength).[3]

Suppose only 10 traits are relevant in a particular RCT, the traits vary little (say, six possible values for each) and the experimental and control groups each contain 10,000 individuals and the sampling was random (i.e. the sampling was such that from the population all possible sets of 10,000 individuals were equally likely to occur). Even this does not warrant the conclusion that the groups are identical in relevant respects. The possible combinations of 10 traits each with six possible values entails, assuming independence, equals 6^{10} (60,466,176). Hence, no two individuals in a population of 60 million are the same.

If the number of relevant factors is 11, the possible combinations equal 362,797,056. The US Census Bureau estimated the population of the United States as of August 8, 2009 to be 307,118,070. Consequently, two groups of 10,000 randomly sampled from the entire population of the United States will not be identical in relevant respects. There are almost always more

[3] Traits determined by multiple genes and subject to environmental influences are known as multifactorial traits. Separating the genetic determinants from the environmental ones is challenging. Also, even with stable environmental factors, understanding the genetic transmission of polygenic traits (those caused by more than one gene) is complex. These traits do not manifest simple Mendelian transmission. Although the transmission is Mendelian, it is a complex process which is the subject of the field of quantitative genetics. Within that field, numerous statistical tools have been developed to deal with this complexity (see: Hartl and Clark, 1989).

than 11 relevant variables (factors) and many more than six possible values for those variables and the population being sampled is always dramatically less than 300,000,000 (often fewer than 50,000). In addition, although the values between individuals can be assumed independent, those in a particular individual are rarely independent of each other – such that altering one will alter one or more of the others – and the specific dynamics are frequently idiosyncratic, which further increases the heterogeneity.

To complicate the situation further, the dynamics involved in multivariable interacting systems, such as those involved in the human endocrine, immune and other such systems are usually chaotic; the trajectory of the system is highly sensitive on initial conditions. Since those initial conditions frequently will be different for different individuals, the trajectory of the systems will vary widely. In short, the homogeneity found in agricultural RCTs is entirely illusory and elusive in clinical medicine; in clinical medicine heterogeneity is ubiquitous.[4] This is why some clinical researchers have endorsed the utility of *n* of 1 trials (i.e. where a single patient is the entire trial).[5]

Fisher's motivation for insisting on randomization was rooted in a requirement for the application of mathematical statistics and the definition of causality he adopted. Randomization is required in order to apply the statistical tools of analysis that underpinned his methods and his causal claims. Given the near impossibility of random sampling in clinical trials, Fisher's methods are not applicable;[6] agricultural trials do not rest on random sampling, they involve the toss of a coin (or some other random binary process) to determine which plot of a pair is the experimental plot. There are, of course, other experimental design methodologies that can be applied in clinical medicine but to successfully avoid the aforementioned challenges their validity must be independent of random sampling. Many other critiques of randomised controlled trials have been offered (see: Cartwright, 2007a; Urbach, 1993; Worrall, 2002 and 2007).[7] For the most part they focus on challenging whether randomization is necessary. I agree with those critiques but here have argued that, whether necessary or not, it is unachievable.

The definition of causality that appears to underlie Fisher's methods is elementary. A cause is that which makes a difference. If two states of affairs at t_1 are identical except for one element E and the states differ at t_2, E can

[4] Of course, there are some cases where the required homogeneity exists. In such cases, the relevant variables are few – usually one or two – and the values are constrained. For example, placing a 10% solution of iodine on the skin will result in discolouration of the contacted area.

[5] See, for example, Guyatt, GH, Keller, JL *et al.* (1990) and Avins, AL, Bent, S, *et al.* (2005).

[6] Fisher, with support from the Rockefeller Foundation, did some work in medicine on blood. He contributed significantly to the understanding of the Rh factor in blood groups.

[7] Many defences have been offered as well (see, for example, Papineau, 1994 and Suppes, 1982). Those defences strike me as inadequate against the criticisms of Cartwright, Urbach and Worrall and do not undermine the criticism I have articulated.

be declared the cause of the difference. In Fisher's agricultural trials, the homogeneity of the states being compared (plots adjacent to each other) virtually ensured identity. By having numerous paired plots (paired states) and randomly selecting which plot from each pair would receive the intervention (E), the validity of the claim that the two plots being compared in each case are identical is certain or near certain. Since, the only difference between compared plots is E, it can with certainty be declared 'the cause' of any difference that arises. The upshot of all this is that Fisher's experimental methods do not provide any basis for discovering or justifying causal claim made on the basis of RCTs in clinical medicine.

2.3 Scientific theories

The conclusion of Section 2.2 is that RCTs in clinical medicine provide no, or at best the weakest possible, support for causal claims.[8] I'll return later to the question, 'For what, if anything, do they provide evidence?' Before doing that, in this section I turn to a positive thesis with respect to clinical medicine – namely, what can and does underwrite causal claims.

Physics and chemistry have a rich toolkit of methods which have teased from nature a large and deep body of knowledge. Noteworthy is the fact that RCTs are not among their methods. There is no shortage of recourse to probability and statistics; indeed they employ the entire domain of mathematical knowledge and techniques (e.g. the infinitesimal calculus, topology, and linear and nonlinear algebra). Physics and chemistry employ probability and statistics in contexts where the phenomena are considered truly random or where, even though a system is held to be deterministic, the current understanding of the system admits of uncertainty. In the latter case, the ultimate quest

[8] It might be thought that, in clinical medicine, RCTs are just testing hypotheses without any clear causal structure. That, in fact, is a role that RCTs could play under the auspices of a causal theory, as I indicate below. That, however, is not how medical epidemiologists view them. With few exceptions, books on clinical epidemiology are quite explicit about the causal goals. Consider, for example, Haynes *et al.*'s book, *Clinical Epidemiology*, in which they claim, 'Our key point is this: RCTs provide the best evidence for causation, so don't give up on doing an RCT to settle a causal issue just because it may be difficult or contentious to do so.' (p. 360). Elwood, in his *Critical Appraisal of Epidemiological Studies and Clinical Trials*, devotes an entire chapter to causality. In the section 'A direct test of causation', he claims, 'If a causal relationship exists, the frequency of the defined outcome will be higher in the group exposed to the causal factor. A study design which uses this approach is the randomized trial; that is, the assignment of the treatment for each subject is made by a random or chance procedure.' (p. 7). Further, Rothman and Greenland in *Modern Epidemiology* also have an entire chapter on causality – one of the most nuanced accounts I have found in epidemiological writings. Uncovering causal relationships is clearly on the minds of epidemiologists who engage in RCTs.

is to diminish the uncertainty; the need to use probability and statistics is unsatisfying, though necessary.

Engineering – an applied endeavour similar to clinical medicine – draws heavily on physics and chemistry. Indeed, much of the confidence we have in the claims, predictions and explanations in engineering rest on the confidence we have in the theories, models and knowledge in physics and chemistry. One fundamental logical feature of engineering reasoning – and reasoning in physics and chemistry – is the use and justification of counterfactual claims (typically expressed as conditional – if, then – claims in which the antecedent – the if part – has not occurred or is not known to be true). For example, the claim, 'If my computer keyboard were in motion relative to me, then it would be shorter in the dimension of travel than it was when stationary relative to me,' is a counterfactual claim since the keyboard is in fact currently stationary relative to me. No physicist, however, would doubt the truth of the claim because its truth rests on Einstein's special theory of relativity; hence, to doubt the truth of the claim is to doubt the validity of that theory.

The reason theories support counterfactuals is that they unify and integrate a large body of knowledge into a connected web.[9] The logical structure of this web is such that explanation and, importantly, prediction rest on a wealth of interrelated knowledge claims. Predictions made on the basis of a theory are possible because the integrated wealth of knowledge claims comprising the theory can be used to justify confidence in the predictions. Prediction is an instance of a counterfactual claim.

Although I hold a view of the structure of theories that understands them as a certain sort of mathematical model – a view I will set out below – the logical empiricist conception of theories as interconnected statements (formalized in first-order predicate logic) is a heuristically useful entry point for uncovering the underlying logic of explanation and prediction. Some statements in a theory are extremely general and cannot be deduced from other statements in the collection; these are the axioms of the theory. In a fully developed theory, all the other statements in the collection can be deduced from the axioms. It is that deductive connectivity that integrates the wealth of knowledge claims; it also justifies confidence in predictions and explanations because they are deductive consequences of the theory.[10] Some claims deduced from the axioms are still very general. The further down the deductive hierarchy one moves, the less general the claims; at the lowest level of generality are claims about specific causes of specific effects (see Figure 2.1 for a stylized schematic diagram).

[9] Quine and Ullian also employ the metaphor of a web in a way analogous to mine (see Quine & Ullian 1970). Kuhn's holistic view of theories also treats them as a web (see Kuhn, 1962).

[10] Deduction is the ideal in deterministic systems and theories describing them but frequently the connections are probabilistic such that the truth of a claim is highly probable based on a collection of other claims in the theory, but not a deductive certainty.

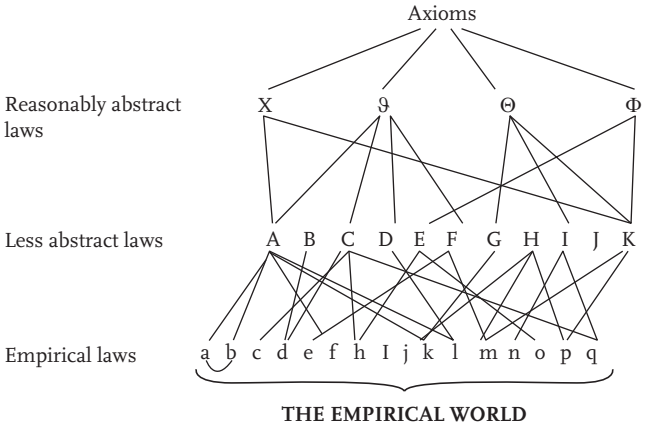

Fig. 2.1 Simple schematic of the deductive web of a theory.

Views differ on how the axioms are generated (discovered). Simplistic empiricism assumes that the first step is the generation of the empirical laws from empirical observation – perhaps by induction, perhaps hypothetico-deductively. Among these empirical laws patterns occur that suggest that a number of empirical laws can be usefully subsumed under a more general claim. Among these more general claims patterns occur that suggest subsets of these more general claims can be subsumed under even more general laws. The process continues until the axioms emerge as the most general claims. Simplistic rationalism assumes the axioms are generated by rational thinking and are subsequently justified by deducing the consequences of accepting the axioms, consequences which are then empirically tested. The history of science suggests that a mixture of these methods is usually involved.

Returning to the claim, 'If my computer keyboard were in motion relative to me, then it would be shorter in the dimension of travel than it was when stationary relative to me', this claim, as noted, is counterfactual. It, however, can be accepted with confidence because it is deducible from the axioms of the special theory of relativity. The degree of confidence, of course, is proportional to the confidence one has in that theory.

The logical empiricist view of theories just sketched assumes that the language of science is first-order predicate logic with identity (symbolic or mathematical logic). The view of theories that I, along with many others,[11] have promoted allows any appropriate mathematical domain to be the language of a theory (e.g. set theory, probability, topology, string theory, and so on). Following Galileo (1623), and three centuries before him Bradwardine (1330),[12] this

[11] See: Suppe (1967, 1972, 1989), Suppes (1957, 1961, 1962, 1967, 1968), Lloyd (1984, 1986, 1988), Thompson (1983, 1986, 1987, 1989, 2007) and van Fraassen (1967, 1969, 1970, 1972).

[12] See: Weisheipl (1967).

view sees mathematics broadly as the language of science. Consequently, it is not a sentential (linguistic) view of the language of science. Theories are not deductively connected statements formalized in symbolic logic but mathematical formulations of dynamical systems. This is still a deductive framework; the deductive structure and techniques of the domain of mathematics used are fully available. Unlike the logical empiricist's view, however, this view understands theories to specify an ontology and the dynamics of a physical system,[13] which is achieved by identifying variables and their range of magnitudes, and specifying, mathematically, the relations among the variables and how the variables change over time (e.g. using transition functions such as $x_{t+1} = f(x_t)$: $f(x_t)$ might be $r\,x_t$ where r is the rate of population growth (births minus deaths) and x is population size). The thesis of this paper does not rest on which view of theories one accepts. Counterfactual claims are deductive consequences of the theory on both views and confidence in them rests on that deducibility.

2.4 Theories, RCTs and causality

A central role of scientific enquiry and scientific knowledge is to answer '**why' questions**, and the perceived importance of uncovering causal connections is motivated by the view that knowledge of cause and effect relationships allows 'why' questions to be answered.[14] Consider the question, 'why, at the onset of the luteal phase in the menstrual cycle, does the level of plasma gonadotropins decrease'. The answer found in *Harrison's Principles of Internal Medicine* (a leading medical resource) is, 'A secondary rise in estrogen **causes** a further gonadotropine suppression' (Braunwald, 2001 p. 2157, emphasis added). Of course, the entire answer to the why question is much more complicated and the additional elements will require the citing additional causes.

In this section, I explore what I argue is a more fundamental element in answering why questions, namely the role of theories. As indicated in

[13] See note 2.

[14] Silvain Bromberger, over 40 years ago, renewed philosophical attention on the importance and role of why questions (Bromberger, 1966). Significant criticisms of Bromberger's specific account have been proffered. I find van Fraassen's early criticism compelling (van Fraassen 1980, pp. 126–130) but do not believe that it diminishes the centrality of why questions; it simply identifies the difficulties with a particular account of the connection between explanation and why questions and the canonical form prescribed by Bromberger. Almost all philosophers of science, and van Fraassen is among this majority, accept that a why question is a request for an explanation and such requests are central to the scientific enterprise. That is all the arguments of this paper require. Of course, van Fraassen and others have put forward a compelling case for the importance of context in determining which theory and/or parts of a theory are relevant to the sought after explanation. I take this as undeniable; a fact that complicates explanation but does not undermine the central arguments of this paper.

previous section, what underwrites confidence in counterfactual claims (and, hence, predictions) is that they can be deduced from relevant parts of a theory. Confidence in the theory comes from the interconnected, and inseparable, nature of the regularities it codifies and the countless predictions deduced using it which have continually been found to be in accordance with the behaviour of the empirical world as we experience it. No isolated claim of a regularity has that robustness and predictions made on the basis of an isolated claim of regularity lack the support and credibility of predictions deduced from regularities embedded in an interconnected web of a vast number and array of regularities. What is true of prediction is equally true of explanation. Although there is no tight symmetry between explanation and prediction, they are two faces of the same logical coin. A robust explanation of an event requires deducing it from a theory, just as a robust prediction requires such a deduction.

To be clear, I am not denying that isolated regularities provide accounts of events; they do. A social worker who asks, 'Why does 8 year old Susan have bruises on her head and shoulder?' may be attempting to determine whether child abuse has occurred. The explanation that Susan fell off a swing the previous day addresses the concern; third-party witnesses make it a compelling explanation. Of course, the social worker has to know that falling off a swing normally results in bruises. For the social worker, this need be no more than an observed regularity – a belief based on the constant conjunction, temporal contiguity and order of the two events. Explanation is pragmatic; the **purpose** of posing a 'why question' determines the relevance of the answer. My claim, as will become clear, is that the purpose of **scientific** research and theorising is to provide an account of observed phenomena, not just a description of them. From the point of view of scientific explanation, the observation that trauma to the skin normally results in a discolouration of the skin in the area of the trauma is simply describing an observation, not explaining it. The goal of scientific[15] research and theorising is to uncover the mechanism that explains the observed phenomenon; in this case, a part of physiological theory.

This goes to the heart of a problem with RCTs. To the extent that they uncover any regularities, they uncover isolated ones unless they can be shown to be among those embedded in, and hence deducible from, a theory. But if such a deduction can be made, it is not the RCT that establishes or justifies acceptance of the regularity; it is the theory. This is not to say that RCTs have no methodological or logical role in scientific enquiry. Fisher's use of them in agriculture is often cast as establishing and justifying causal assertions.

[15] I, of course, am using 'scientific' in a narrower sense that some others might. I think my use accords with standard usage in philosophy of science. Nonetheless, little hangs on this. If forced to broaden the scope of the phrase 'scientific explanation', I would use 'robust scientific explanation' in its place.

But that is the wrong way to view things. RCTs divorced from a theory provide, at best, knowledge of isolated regularities (for Fisher a cause and effect relationship) and, as already argued, isolated regularities lack the robustness required in providing compelling predictions or explanations. What RCTs, used as Fisher used them, can do is provide a method for testing predictions made using a theory – in effect, they are a way of empirically testing a theory. In this way, they confirm or call into question some feature of the theory, such as one or more of its axioms, the validity of a particular deduction from it or some interpretation of its ontology or dynamics, and the like. This is an important role but one that can **only** be played in the context of a theory. Moreover, this role is not an explanatory one.[16] Furthermore, in the absence of a theory, determining whether and in what way an RCT should be conducted is doomed to failure.[17]

An example from immunology illustrates the distinction I am drawing between an isolated result from an RCT (or several RCTs that produce similar results) and a robust explanation of the result that embeds it in a theory. In the eighteenth century, protection from smallpox was discovered to occur in some individuals after inoculation with the dried material from a smallpox pustule. Regrettably, about three in 100 people developed a severe case of smallpox and died. Edward Jenner, in 1796, discovered that inoculation with material from cowpox pustules (the bovine form of smallpox) also conferred protection without causing severe cases. He called this inoculation process vaccination (from *vacca* –Latin for cow: also the origin of vaccinia for the virus that causes cowpox); Pasteur (honouring Jenner) extended this term to cover all inoculations which provide protection from infectious agents. Jenner had no knowledge of the infectious agent; he only knew that protection from smallpox followed vaccination with cowpox, with only a small number of individuals developing serious disease. The vaccine was modified during the following century and a half (e.g. attenuated versions of the smallpox virus were developed).

Jenner's experiment to demonstrate the efficacy of inoculation with cowpox falls significantly short of the 'proof' required today and would completely fail an ethics test. Jenner inoculated an eight-year old boy, James Phipps, with cowpox material. He waited six weeks and inoculated him with fluid extracted from an active smallpox pustule. Phipps did not contract smallpox. On the basis of this 'experiment', he published his success in 1798. As one

[16] Nancy Cartwright, with somewhat different arguments and purposes, has made a similar point in her compelling and insightful recent book, *Hunting Causes and Using Them* (Cartwright, 2007), as also have numerous philosophers over the last 50 or so years (see also Cartwright, 2008).

[17] This is a point elegantly made by Hempel (1966, pp. 10–18). His example and argument are entirely independent of his logical empiricist philosophy; it applies equally to other views of theories and their role in science.

would expect, smallpox vaccines approved in the latter half of the twentieth century were subjected to RCTs. Clearly, an RCT would have provided a higher level of confidence in the efficacy of Jenner's vaccine. It would not, however, have added anything to Jenner's description of the connection of the events. RCTs on more recent smallpox vaccines provide evidence that vaccination is followed by protection against smallpox – knowledge that is, without question, valuable in clinical medicine[18] – but an explanation of why there is a connection is not provided by an RCT.

What does provide an explanation is immunological theory; it provides an account of how the vaccine results in immunity to the *variola major* virus (the virus causing smallpox); and theories in virology and physiology provide an account of how the virus causes the clinical manifestations of smallpox. A comprehensive account of the explanation is complex and more technical than appropriate for this paper but the skeleton can be easily provided. The virus in Jenner's vaccine has the same antigenic determinants (epitopes) as the variola virus but is not a viable pathogen in humans. Lymphocytes (a kind of white blood cell) are produced in the bone marrow. A common lymphoid progenitor gives rise to lymphocytes. There are two major kinds: B lymphocytes (which mature in the bone marrow – hence B) and T lymphocytes (which mature in the thymous – hence T). These are more commonly known as B cells and T cells. Lymphocytes recognise specific sites that are present on antigens – foreign material. A large variety of site-specific T lymphocytes are produced, each recognizing a different epitope. The presence of the vaccine virus is detected by a specific lymphocyte whose receptor matches the virus' epitope; that detection results in the production of a large number of lymphocytes that are site-specific for the virus. Through a complicated biochemical process, the production of armed effector T cells with that site-specific receptor is initiated. These effector T cells inactivate the virus by binding to the epitope. As part of the process, memory T cells and B cells are produced; these provide the observed long-term protection. The same process can be initiated by inoculation with attenuated variola (smallpox) virus.

What makes this a robust explanation of the observed phenomena is the rich body of generalizations on which it draws and the rich ontology involved (hematopoietic stem cells, neutrophils, B cells, T cells, antibodies, basophils, and so on). The deductive network of generalizations at a variety of levels of generality integrates this single-case connection of events (vaccine and protection) in a large framework. A framework that also explains why lymphocytes do not bind to the bodies own tissues, and why the major histocompatability complex (MHC) of genes is important to the production of armed effector

[18] Smallpox vaccination has resulted in one of the great successes of clinical medicine. The last reported case of smallpox was in Somalia in December 1977. On 9 December 1997, the World Health Organization declared that smallpox had been eradicated.

T cells, and why B cells bearing surface CD5 express a distinctive repertoire, and why HIV produces an autoimmune response, and so on and so on. The connection of vaccine and protection is imbedded in this complex dynamical system; a system by means of which we can provide a rich multilayered explanation of the observed connection of events.

Importantly, immunological theory explains the heterogeneity of individuals and explains the heterogeneity of responses to interventions. The explanations will appeal, for example, to genetic differences, such as differences in the MHC group of genes, to compromises to the system, such as low leukocyte counts, to deficiencies in critical precursor elements, such as cytokines, and so on. This explanatory power of heterogeneity is a feature of all theories in medicine.

By contrast, RCTs, independent of this theoretical framework, focus on an isolated connection of events. Even a meta-analysis, which examines and analyses numerous RCTs, focuses on an isolated connection of events. A meta-analysis may provide even stronger evidence that there is a connection but it stills fails to explain why there is a connection. And, the heterogeneity of individuals and their responses to interventions bedevils RCTs divorced from a theory.

Returning now to Fisher to further support my thesis, as noted, for the most part, Fisher's method of experimental design was focused on agricultural research. He did, however, with support from the Rockefeller Foundation, do some work in medicine on blood groups during the period 1935–1943. This work contributed significantly to the early understanding of the Rh factor. It, however, drew heavily on population genetics and evolution, both are robust theories.[19] What Fisher demonstrated in 1943 (see Fisher 1943 and 1944), using the theory of population genetics, was the role that three linked loci with specific allelic combinations could play in the explanation of the puzzling experimental results with the Rh factor. He also predicted, again on the basis of population genetical theory, the existence of antibodies not known to exist at that point. Within the next five years his prediction and explanation received independent empirical support – further confirming the theories. Consequently, Fisher's work in clinical medicine, far from demonstrating the value of RCTs in that domain, elegantly demonstrated the value of a robust theory in providing explanations and making predictions.

Fisher's work on population genetics and evolution and his use of them in medical explanation and prediction makes clear that probability and statistics play an important role in science, outside of RCTs. The domain of mathematics employed as the language of his genetical theory of natural selection is probability and statistics. Using probability and statistics in this way – i.e. as

[19] Fisher contributed significantly to their development (Fisher, 1930).

the language of theory – is common in physics, chemistry and biology; its use in RCTs is not.[20]

What is being questioned here is the appropriateness of the use of probability and statistics in RCTs in medicine. In agriculture, many of the presuppositions on which a legitimate use of probability and statistics are based are met; this is the case in medicine. In medicine, randomization is almost always gerrymandered (sampling is not from the entire relevant population, some individuals assigned to a sample are removed after the fact, samples are adjusted to eliminate relevant differences observed after sampling or known to be likely from past experience – difference in age profile or imbalanced gender, for example – and so on). In addition, the assumption of homogeneity that is reasonably robust in Fisher's agricultural work is absent in medicine,[21] which in part accounts for 'side effects' which are often more prevalent that the target effect, the heterogeneity of outcomes[22] and the constant publication of contradictory findings about the same intervention. The heterogeneity in the population and in outcomes undermines any chance of justifying causal claims. In Fisher's agricultural trials, justifying a causal claim on the basis of the trial is plausible. The problem in this case, as I have argued, is that the causal claim is isolated and, hence, cannot provide a basis for explanation; for that a theory is required. Unlike Fisher's agricultural trials, medical trials do not come remotely close to even justifying the assertion of a causal claim.

In agriculture, randomization is restricted to choosing one of a matched pair of plots – not as in medicine to sampling from a population; paired plots can with a high level of confidence be assumed homogeneous for all relevant factors – contrary to the situation in medicine; and the multiple match-plots which are part of each trial provide replication within the trial – in medicine, replication requires new trials which will be few in number and almost certainly dissimilar to the original trial in important ways.

[20] To avoid any confusion, let me be clear that, as the forgoing use by Fisher makes abundantly clear, the importance or power of probability and statistics in science is not in question. Fisher's use of that domain in population genetics, evolution and medicine parallels its use in physics (e.g. statistical mechanics (a deterministic sphere) and quantum mechanics (an indeterministic sphere)), chemistry (e.g.) and biology (e.g. population genetics). Patrick Suppes used it in a compelling way to develop a probabilistic theory of causality (Suppes, 1970). It worth noting in passing that Suppes is quite clear about the role and importance of theory in his account, 'The analysis of causes and their identification must always be relative to a conceptual framework [what I take a currently accepted theory to be], and there is no successful argument apparently that can show that a particular conceptual framework represents some ultimate and correct view about the structure of the world' (Suppes, 1970, pp. 90–91). There are also a host of other ways in which probability is used in science – from determining goodness of fit between predictions deduced from a theory and the experimental data obtained to describing the distribution of chance events.

[21] See: Upshur (2005).

[22] See: Kravitz *et al*. (2004).

2.5 Causality, theories and an eliminative thesis

Causality has had a rough couple of millennia. Aristotle identified four causes (efficient, formal, final and material). Today 'cause' is only associated with his efficient cause; the other three are held in various states of derision. Even efficient cause has been, and still is, under constant attack. It was, for example, pummelled by David Hume and outright rejected by Bertrand Russell. Responses and counter-responses abound. Patrick Suppes opens his excellent *A Probabilistic Theory of Causality* by quoting the relevant passages from Russell's, 'On the Notion of Cause'; he then argues that Russell's position is a relic of a superseded period of physical science – a period in which 'the fundamental physical phenomena in question were felt to be much better understood at a fundamental level than they are today'. Since, I shall provide a neo-Russellian account repeating the quotation here provides an appropriate starting point.

> All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word 'Cause' never occurs. . . . 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. . . . The principle 'same cause, same effect,' which philosophers imagine to be vital to science, is therefore utterly otiose. As soon as the antecedents have been given sufficiently fully to enable the consequent to be calculated with some exactitude, the antecedents have become so complicated that it is very unlikely they will ever recur. Hence, if this were the principle involved, science would remain utterly sterile. . . . No doubt the reason why the old 'law of causality' has so long continued to pervade the books of philosophers is simply that the idea of a function is unfamiliar to most of them, and therefore they seek an unduly simple statement. There is no question of repetitions of the 'same' cause producing the 'same' effect; it is not in any sameness of causes and effects that the constancy of scientific laws consists, but in 'sameness of relations'. And even 'sameness of relations' is too simple a phrase; 'sameness of differential equations' is the only correct phrase.

Suppes is correct to point out that the natural sciences at the time he was writing (1970) were more conceptually and theoretically complex than in the period Russell was writing (1910–15); the full impact of Einsteinian relativity (his special theory of relativity was published in 1905 and his general theory of relativity was presentation to the Prussian Academy of Science in 1915) had not occurred and quantum theory only began to coalesce in the 1920s. Since Suppes' observation, the natural sciences have been influenced by chaotic dynamical systems, fractal mathematics and computer simulation to mention but a few important factors. These changes make Suppes' observation more apt. Notwithstanding, however, this correct observation of Suppes', I contend that Russell uncovered the kernel of a profound reinterpretation of causality.

The essence of the argument can be sketched by reflecting on the history of teleology in physics and biology. In the period from Galileo to Newton, the role of teleological accounts of phenomena shrank in physics and astronomy. After Newton, its role was miniscule to non-existent – bursts of attempted resuscitations were unsuccessful. There are cases where the language used appears to invoke a teleological account. For example, it might be said that a particular missile is 'seeking' a fighter aircraft. The missile changes trajectory as the target moves. That phenomenon is a function of an internal positive/negative feedback mechanism; the language of 'seeking' is simply a shorthand expression that can readily be replaced by the mechanistic one. No physicist or engineer – indeed no moderately educated person – would really believe that the missile was 'seeking' the aircraft and was directing its behaviour to achieving that goal. It is simply behaving in accordance with its internal structure and program. If anything can be identified as a goal, it is a humanly constructed goal of incapacitating the aircraft within a larger military goal. Whether this 'goal' is irreducible to a mechanism is complex and irrelevant to the goal-directed language of the physical entity and its behaviour; the physical entity and its behaviour have no intrinsic goals.

In biology, teleological accounts of phenomena were alive and well until the second half of the nineteenth century; the publication of Darwin's *Origin of Species* began the slow decline in the use of teleological accounts. It is not that the use of teleological language was purged from biology. Indeed, today it is easy to find such language in biological books and articles – especially those dealing with the behaviour of organisms. What changed with Darwin was how this language was understood.

When a biologist remarks, 'Hymenoptera perform this dance **in order to** communicate the direction and distance of a food source to other workers,' there is no attribution of intentions. The use of 'in order to' is shorthand for a mechanistic understanding. The behaviour has a genetic basis and is, therefore, biologically programmed and heritable, and the genes responsible for the behaviour have become ubiquitous in that species because they enhance reproductive success. There is no goal of communication, in any teleological sense – though information is in fact conveyed. To the extent anything can be considered a goal, it is the reproductive success of the individual organisms.

Although Suppes' (1970) observation that causal language is pervasive in modern sciences remains true today, I contend that it is, nonetheless, like the use of teleological language; it is a shorthand expression, which owes any meaning and validity to the existence of models and theories. It can, and often should, be eliminated in favour of a mechanistic theoretical account. The claim 'A caused B' is shorthand for 'The claim, whenever the system is in state A, the next state of the system will be B (either always or with $\Pr(x)$) can be deduced from a currently accepted and well-confirmed dynamical theory.'

On this interpretation of 'cause', even in agricultural trials, RCTs do not justify causal claims; only a theory can do that. In addition, this interpretation goes to the heart of a diagnosis of the problem with RCTs in medicine. The problem identified above is that in the absence of a theory, RCTs do not provide explanations or allow predictions. That is a hefty shortcoming, and were not so much at stake it would make the claim that RCTs are a gold standard risible. They do not provide explanations and predictions because the results, unless connected in ways I have described to a theory, stand isolated; an isolation, this reinterpretation of 'cause' suggests, renders causal claims made on the basis of RCTs vacuous.

In physics, chemistry and biology, Russell overstated the case with his claim, '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.' In those contexts, using the shorthand language of causality almost never results in researchers failing to grasp the importance and role of a theory. Those researchers quite naturally – almost unconsciously – appeal to theories in making predications and providing explanations. Russell's claim, however, has considerable validity in clinical medicine that focuses on RCTs. An impoverished interpretation of causality that divorces it from theories results in significant methodological, epistemological and logical harms. In turn, these result in the harm of suspect findings and claims, and poorly or improperly understood interventions. At the heart of the harm is the undermining of the validity of explanations, predictions and clinical treatments; and that is far from a trivial harm.

The conclusion to draw from all this is not that RCTs and the results of RCTs have no role in medicine; they do. The appropriate conclusion is that their role is dependent on being integrated into (indeed subservient to) a theory. Fortunately, in spite of the emphasis on RCTs, robust theories, that can be used to ground RCTs, abound in medicine; from immunology through to physiology and endocrinology to neurosciences robust theories are found. It is these that have provided the solid, lasting basis for medical explanation, prediction, diagnosis and treatment. Consider, for example, the compendium on immunobiology by Charles Janeway Jr.[23] or the text on medical genetics by Margaret Thompson *et al.*[24] Their entire treatment of their subject is experimental and theoretical, and, most notably, RCTs play no role in the evidence, explanations and predictions provided.[25]

[23] Janeway (1997).

[24] Thompson *et al.* (1996).

[25] The same is true in many other research fields in medicine as a perusal of medical texts in medical genetics, human physiology, neurosciences will reveal.

To appropriate Dobzansky's famous claim that, 'Nothing in biology makes sense except in the light of evolution',[26] 'nothing makes sense in medicine except in the light of a theory'.

Acknowledgements

I am grateful to two reviewers for providing suggestions for strengthening the expositions and arguments of this paper.

References

Avins, A.L., Bent, S., *et al.* (2005). Use of an embedded N-of-1 trial to improve adherence and increase information from a clinical study. *Contemporary Clinical Trials* 26 (3): 397–401.

Bradwardine, T. (*circa* 1330). *Tractatus de Continuo*. [A Latin text based on the three extant manuscripts can be found in J.E. Murdoch (1957) Geometry and the continuum in the fourteenth century: a philosophical analysis of the Thomas Bradwardine's Tractatus de continuo (PhD dissertation: University of Wisconsin)].

Braunwald, Eugene *et al.* (eds.) (2001). *Harrison's Principles of Internal Medicine*, 15th edition, New York: McGraw-Hill.

Bromberger, S. (1966). Why-questions, in R.G. Colodny (ed.) *Mind and the Cosmos*. Pittsburgh: University of Pittsburgh Press, pp. 86–108.

Cartwright, N. (2007a). Are RCTs the gold standard? *BioSocieties*, 2: 11–20.

Cartwright, N. (2007b). *Hunting Causes and Using Them*. New York: Cambridge University Press.

Cartwright, N. (with Andrew Goldfinch and Jeremy Howick) (2008). Evidence-based policy: Where is our theory of evidence? in A. Beckermann, H. Tetens and S. Walter (eds), *Philosophy: Foundations and Applications*. Paderborn: Mentis-Verlag.

Dobzhansky, Theodosius (1964). Biology, molecular and organismic. *American Zoologist* 4: 443–452

Dobzhansky, Theodosius (1973). 'Nothing in Biology Maths sense except in the light of evolution *American Biology Teacher*, 35, No. 3(March): 125–12.

Elwood, M. (1998). *Critical Appraisal of Epidemiological Studies and Clinical Trials* 2nd edition. Oxford: Oxford University Press.

Fisher, R.N. (1930). *The Genetical Theory of Natural Selection*. London: Oxford University Press.

Fisher, R.A. (1935). *The Design of Experiments*. Edinburgh: Oliver & Boyd.

Fisher R.A. and Fraser Roberts, J.A. (1943). A sex difference in blood-group frequencies. *Nature*, 151: 640–641.

[26] Dobzhansky (1964) p. 449, see also, Dobzhansky (1973). Dobzhansky meant by 'evolution' both the fact that it occurred and, most importantly for him, the modern synthetic theory of evolution.

Fisher, R.A., Race, R.R. and Taylor, G.L. (1944). Mutation and the Rhesus reaction. *Nature*, 153: 106.

Galileo, G. (1623). *Il saggiatori* (reprinted in English translation in C.D. O'Malley and S. Drake (eds) *Controversy on the Comets of 1618*, Philadelphia: 1960.

Guyatt, G.H., Keller, J.L., Jaeschke, R., Rosenbloom, D., Adachi, J.D., Newhouse, M.T. (1990). The *n*-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. *Annals of Internal Medicine* 112 (4): 293–299.

Hartl, D.L. and Clark, A.G. (1989). *Principles of Population Genetics*. Sunderland, Massachusetts: Sinauer Associates, Inc.

Haynes, R.B., Sackett, D.L., *et al.* (2006). *Clinical Epidemiology: How to Do Clinical Practice Research* (3rd edition). Philadelphia, PA: Lippincott Williams and Wilkins.

Hempel, C.G. (1967). *The Philosophy of Natural Science*. Englewood Cliffs: Prentice Hall.

Janeway, Jr., C.A., Travers, P. *et al.* (1997). *Immunobiology: The Immune System in Health and Disease* (3rd edition). New York: Garland Publishing, Inc.

Kuhn, Thomas S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kravitz, Richard L. *et al.* (2004). Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly* 82: 661–687.

Lind, J. (1753). *A Treatise on Scurvy*. Edinburgh: Sand, Murray and Cochan.

Quine, W.V. and Ullian, J.S. (1970). *The Web of Belief*. New York: Random House.

Rothman, K.J. and Greenland, S. (eds.) (1998). *Modern Epidemiology* (2nd edition). Philadelphia, PA: Lippincott Williams and Wilkins.

Suppe, F. (1967). *On the Meaning and Use of Models in Mathematics and the Exact Sciences*. Ann Arbor: University Microfilms International (Ph D dissertation).

Suppe, F. (1972). What's wrong with the received view on the structure of scientific theories? *Philosophy of Science* 39: 1–19.

Suppe, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. Urbana: University of Illinois Press.

Suppes, P. (1957). *Introduction to Logic*. Princeton: Van Nostrand.

Suppes, P. (1961). A Comparison of the meaning and Uses of Models in Mathematics and the Empirical Sciences, in Hans Freudenthal (ed.) The concept and the Role of th Model in Mathematics and Natural and social sciences Dordrecht: D, Reidel, (1961) pp. 163–177.

Suppes, P. (1962). Models of data, in E. Nagel, P. Suppes and A. Tarski (eds.) *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*. Stanford: Stanford University Press, pp. 252–261.

Suppes, P. (1967). What is a scientific theory? in S. Morgenbesser (ed.) *Philosophy of Science Today*. New York: Basic Books, pp. 55–67.

Suppes, P. (1968). The desirability of formalization in science, *Journal of Philosophy* 65: 651–664.

Thompson, M.W., McInnis, R.R., *et al.* (1966). *Genetics in Medicine* (5th edition). Philadelphia, Pennsylvania: W. B. Saunders Company.

Thompson, P. (1983). The structure of evolutionary theory: A semantic approach. *Studies in History and Philosophy of Science* 14: 215–229.

Thompson, P. (1986). The interaction of theories and the semantic conception of evolutionary theory. *Philosophica* 37: 73–86.

Thompson, P. (1987). A defence of the semantic conception of evolutionary theory. *Biology and Philosophy* 2: 26–32.

Thompson, P. (1989). *The Structure of Biological Theories*. Albany: State University of New York Press.

Thompson, P. (2007). Formalisations of evolutionary biology, in M. Matthen and C. Stephens (eds.) *Philosophy of Biology*. Amsterdam: Elsevier, pp. 485–523.

Upshur, Ross E.G. (2005). Looking for rules in a world of exceptions: Reflections on evidence-based practice. *Perspectives in Biology and Medicine* 48: 477–489.

Urbach, P. (1993). The value of randomization and control in clinical trials. *Statistical Medicine*, 12: 1421–1431.

van Fraassen, B.C. (1967). Meaning relations among predicates. *Nous* 1: 161–179.

van Fraassen, B.C. (1969). Meaning relations and modalities. *Nous* 3: 155–167.

van Fraassen, B.C. (1970). On the extension of beth's semantics of physical theories. *Philosophy of Science*, 37: 325.

van Fraassen, B.C. (1972). A formal approach to philosophy of science, in R.E. Colodny (ed.) *Paradigms and Paradoxes*, Pittsburgh: The University of Pittsburgh Press.

van Fraassen, B.C. (1980). *The Scientific Image*. Oxford: Clarendon Press.

Weisheipl, J.A. (1967). Galileo and his precursors, in E. McMullin (ed.) *Galileo: Man of Science*. New York: Basic Books.

Worrall, J. (2002). What evidence in evidence-based medicine? *Philosophy of Science* 69: S316–S330.

Worrall, J. (2007). Why there's no cause to randomize, *British Journal for the Philosophy of Science* 58 (3): 451–488.

·3·

Inferring causation in epidemiology: Mechanisms, black boxes, and contrasts

Alex Broadbent

Abstract

This chapter explores the idea that causal inference is warranted if and only if the mechanism underlying the inferred causal association is identified. This *mechanistic stance* is discernible in the epidemiological literature, and in the strategies adopted by epidemiologists seeking to establish causal hypotheses. But the exact opposite methodology is also discernible, the *black box stance*, which asserts that epidemiologists can and should make causal inferences on the basis of their evidence, without worrying about the mechanisms that might underlie their hypotheses. I argue that the mechanistic stance is indeed a bad methodology for causal inference. However, I detach and defend a mechanistic interpretation of causal generalisations in epidemiology as existence claims about underlying mechanisms.

3.1 Causal hypotheses in epidemiology

What does it take to establish a causal hypothesis in epidemiology? What standards need to be met? Or, if establishment comes in degrees, degrees of what?

The most obvious aspect of this problem concerns inferring causation in a particular study. A study reveals a statistical association between smoking and lung cancer, or a certain gene and obesity. Statistical analysis reveals a low p-value – a measure of the chance that the association is due to chance. Study design controls for confounding variables (what philosophers would call common causes of the putative cause and effect). Can it be inferred that, for *this* group, a causal relationship exists between smoking and lung cancer, or having that gene and obesity?

Oddly enough, this is not a question that epidemiologists like to answer. A single study would not normally be considered a sufficient basis for a causal inference. Replication is a guiding epidemiological principle. From a

methodological point of view this is extremely interesting. Epidemiologists' credence in a causal hypothesis about Study Group A increases when the effect is replicated in Study Group B. Explaining (or, I suppose, refuting) this attitude is a central task for any methodological analysis.

A second difficulty concerns the inference from a study, or a collection of studies, to a wider population. Epidemiologists are centrally concerned with extrapolating from the people they study to people they have not studied. Replication is important here too, because one way to argue that differences between the population studied and the target population are causally irrelevant is to replicate the study among people who are drawn from the target population. However, replication cannot solve the problem of generalisation. Often the study group will *already* be drawn from the target population: for example, when generalizing from the Whitehall studies to the population of Britain.[1] Differences between those studied and those not studied will always remain; the difficulty is working out when these differences make a difference. On other occasions, studies on a subset of the target population may be impractical: for an obvious example, consider future populations. Quite generally, a central purpose of epidemiology is to get more for less: to learn something not only about those who *have* been studied, but about those who *have not*.

Epidemiologists make efforts to be precise about the scope of their claims, by explicitly stating whether they are intended to apply to the group studied, or to a wider population, and if the latter what conditions the wider population are to meet. For example, instead of saying 'genetic influences cause paediatric obesity', they might say:

> Genetic influences on BMI and abdominal adiposity are high in children born since the onset of the pediatric obesity epidemic.
>
> (Wardle *et al.* 2008, p. 398)

However, even this is an incomplete specification. The children studied were British, but the obesity epidemic affects Europe and America too. Are these results evidence for high genetic influences on BMI and adiposity in children in Britain only, or in Europe and America too? The incompleteness of the specification is not necessarily a failing of the authors of the study. It reflects the genuine difficulty of deciding how to generalize.

Note that replicating the study in European or American children is a way to *avoid* the question, not to answer it. Replication cannot tell us whether a generalisation from this study to American children would be *warranted*, only whether it would be *correct*. In circumstances where we can replicate, that may

[1] The Whitehall studies concerned various aspects of social status and health among British civil servants (Marmot 2006). The aim was to identify social determinants of health, especially the role of purely social differences in explaining differences in health, by studying a group of people whose absolute wealth was such that their basic biological needs were met.

be the best strategy; but for reasons I have already given, we cannot universally substitute replication for generalisation.

Thirdly, there is a difficulty interpreting general claims, even when their scope is fixed. A great deal of philosophical attention has been directed towards singular causal claims, such as 'Jones' smoking caused his lung cancer'. But epidemiologists are almost exclusively concerned with general claims, such as 'smoking causes lung cancer'. Does the latter express a relationship between smoking and lung cancer, or is it a generalisation over individual causal relationships – along the lines of '*in X% of cases*, smoking causes lung cancer'?

This difficulty is a relatively familiar one to philosophers, but it is perhaps not the most pressing one for epidemiologists. In practice, epidemiological hypotheses are explicitly exception-ridden. Accordingly they are framed not as universal generalisations, but as measures of the *influence* of one factor on an outcome, or measures of the *strength* of an association, or of the *proportion* of an effect that is due to a particular factor or group of factors. These sorts of claims raise what is fundamentally the same problem, but in a slightly different way. For example, saying 'Genetic influences . . . are high' makes it clear that the generalisation is not *exceptionless*. But it still does not make clear exactly how the degree of influence is to be interpreted. Is the claim that, in each individual, the genetic influence is high? – This would amount to a universal generalisation attributing a certain genetic influence to each individual. Such an interpretation is hard to make sense of on either the effect side or the cause side. On the effect side, obesity might be absent in some of the individuals studied. Genetic factors cannot then be said to influence it. Switching from a qualitative property (obesity) to a quantitative one (such as bodyweight) will not always be straightforward: the absence of effects such as lung cancer, diabetes, and suicide are hard to interpret as zero degrees on any quantitative scale. Similarly on the cause side, it makes little sense to attribute some degree of influence to a factor that is absent. This is not clear in the example I have picked, since 'genetic influences' are always present in people, but it is obvious when we consider single-gene conditions. When somebody lacks the gene but has the trait in question, it makes no sense to attribute the trait's presence to the influence of the absent gene in *any* degree.

Another interpretation would see measures of influence, proportion, strength of association, and similar, as measures of the *proportion of cases in which a factor is causal*. (This interpretation is akin to the generalisation-over-singular-causation view of universal causal generalisations.) This view is easy enough to make sense of, but there is a case that it reflects metaphysical commitments rather than epidemiological evidence. Take a measure such as heritability, which is the proportion of a given trait in a given population that is due to genetic factors. The idea that causes are either present or absent,

strictly speaking, and the view that we cannot quantify the contribution of a *particular* cause to a *particular* effect, are widespread among philosophers from John Stuart Mill to David Lewis (Mill 1843; Lewis 1973). On such a view, heritability expresses the proportion of the *population* in which the trait is caused by genetic factors – but in each *individual*, the trait either is or is not caused by genetic factors. But this interpretation is not stable, because on this metaphysical picture, causation is not exclusive. Saying that a trait is caused by genetic factors in an individual is compatible with saying that it is caused by non-genetic factors: events have many causes. In the study I have been using as an example, the aim is to measure the *contribution* of genetic influence to obesity in a population. To interpret this as a claim about the proportion of individuals in whom genetic factors cause obesity would be bizarre, since genetic factors are part of the causal history in 100% of cases of obesity. Indeed, every trait is both 100% genetic and 100% environmental, on this interpretation (Rothman and Greenland 2005, S146). Better, we could see it as a claim about the proportion of individuals in whom genetic factors *make the difference* between being obese and not. This interpretation might be made to work; but that would be no trivial philosophical achievement. The interpretation of heritability is a topic of considerable dispute (e.g. Schonemann 1997; Sesardic 2005).[2]

Two lines of response to this bundle of difficulties may be discerned in the contemporary methodological-epidemiological literature. These lines of response are in tension. One is the *mechanistic stance*: the view that causal inference in epidemiology aims at discovering mechanisms: that discovering mechanisms is necessary and sufficient for establishing a causal hypothesis. The other is the *black box stance*: the view that epidemiology is primarily concerned with statistical analysis of associations, and only incidentally concerned with uncovering mechanisms. In Sections 3.2 and 3.3 I will describe and evaluate each stance, and in Section 3.4 I will propose a resolution.

My terms 'stance', 'line of thought', and similar are intended to avoid commitment on the question of whether any actual epidemiologist wholeheartedly asserts any of the views discussed. I rather doubt that any does. Nonetheless, these are not straw men: these stances are evident in the methodological writings of actual epidemiologists, and there is value in seeking to draw them out into the light for explicit evaluation, even though – indeed, partly because –

[2] Heritability is not to be confused with heredity. An individual can inherit a trait from her parents, for example her eye colour. The disposition to develop blue eyes given a certain broadly congenial environment is hereditary; favourite colour probably is not. The heritability of a trait is defined with reference to a population, and makes no sense applied to an individual. It is meant to be a measure of the *relative* contribution of genes and environment, and is not fixed for a given trait. Hence Wardle's interest in showing that obesity is *still* highly heritable despite an increase in the availability of calories.

nobody would endorse these views when stated explicitly and taken to their logical conclusions.

3.2 Mechanisms

There has been a surge of interest in mechanisms in recent philosophy of science. One well-known definition is this:

> A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalisations.
>
> (Glennan 2002, S344)

The word 'mechanism' is less common in epidemiology than it is in some other biological sciences. Nevertheless, it would be interesting if a similar idea could be identified in actual epidemiological practice.

I think it that it can. Epidemiologists, like neuroscientists, use the term 'mechanism', and they do so in revealing ways, as I shall argue. But in addition, I suggest that it is plausible to take the common epidemiological phrase 'causal pathway' as referring to a mechanism. Perhaps a causal pathway is not exactly the same thing as a mechanism: for one thing, it may be longer, and include several 'mechanisms' in the sense intended by Glennan. Nevertheless, causal pathways probably do meet the proffered criterion for mechanisms, since they typically will be postulated to explain a 'behaviour', and will plausibly constitute a 'complex system' whose parts interact according to 'direct, invariant, change-relating generalisations'. Moreover, identifying a mechanism in neuroscience, and identifying a mechanism in epidemiology, satisfy the same goal: they both explain *how something works*. For these reasons, it is plausible to see the search for causal pathways in epidemiology as a search for what Glennan and others call 'mechanisms', even if the causal pathways identified in epidemiology are not terribly similar to the mechanisms in neuroscience.

The notion of mechanism neatly captures one methodological story that may be discerned in the epidemiological literature. That story has two parts. Initially, associations are identified between variables and health outcomes. By a sort of process of elimination, it is established that it is very unlikely that the association is due to chance, or to confounding variables (common causes, to philosophers) or other biases. A variable for which such an association has been established is called a *risk factor*. This first part of the process provides a good reason to think that the risk factor is causal, but a second stage is required for a conclusive case. The *mechanism* for the operation of the risk factor must be identified. Perhaps not immediately; but if in the fullness of time no mechanism is identified, the credibility of the hypothesis will suffer.

Moreover establishing a mechanism is sufficient for proving that a causal hypothesis is correct: showing *how* A actually causes B is conclusive proof that A *does* cause B.[3] The *mechanistic stance* is the methodological position that identifying the underlying mechanism is both necessary and sufficient for warranted inference to a causal hypothesis.

As a description of contemporary epidemiological methodology, the mechanistic stance is appealing. It has plenty of illustrations. Genetic epidemiology is an extremely good fit. Consider this extract from the introduction of another high profile clinical study in the genetics of obesity:

> The genetic contribution to body weight has been established through family studies, investigation of parent–offspring relationships, and the study of twins and adopted children . . . As is the case for height . . . environmentally driven changes in body weight in the population occur against a background of susceptibility to weight gain that is determined by genetic factors. Thus, genetic approaches can be applied to understand both the molecular and physiological mechanisms involved in human obesity.
>
> (Farooqi and O'Rahilly 2006, p. 710)

The rationale appears to follow exactly the lines of the mechanistic stance I have described. There is considerable evidence for a causal link between genetics and weight; the next step is to understand the mechanisms involved.

A similar sentiment is evident at what might be seen as the other end of the epidemiological spectrum, among those who work on social determinants of health. They, too, see it as crucial to identify plausible pathways for the determinants they identify. Here is an extract from another introductory rationale, this time from a chapter by two leading proponents of social determinants of health research:

> Psychosocial factors and their influences on health are active areas of research . . . There is now enough evidence to suggest that this is an important field for those concerned with improving public health . . . Plausible mechanisms linking psychosocial factors to health are described in the first half of this chapter. We then look to the evidence from both human and animal literature to illustrate the ways in which social organisation can influence our biology and, therefore, the health of individuals and populations.
>
> (Brunner and Marmot 2006, p. 8)

This passage mixes the mechanistic stance as a purely methodological stance with the natural companion view that identifying mechanisms is a good way to improve public health. Setting that aside, the mechanistic stance is clearly discernible. There is already evidence that psychosocial factors influence health; the purpose of the chapter is to work out how, by postulating 'plausible mechanisms' and presenting 'evidence from both human and animal litera-

[3] Showing that a given mechanism is how A *actually* causes B must not, of course, be confused with exhibiting a mechanism by which A *could* cause B. The entailment goes through only when what has been shown is how A *actually* causes B.

ture' in support. Whether or not this is the authors' intention, they certainly give the impression that providing a mechanism a key element in the case for the hypothesis that social status is a determinant of health.

In each of these cases, the identification of mechanisms is seen as important: important enough to devote an entire paper to. Why? It *could* be simple scientific curiosity: the desire to find out how things work, just for the sake of it. But I think there is more to it than that. Identifying mechanisms is presented as important not only to complete the scientific picture, but also to seal the case for existing causal hypotheses. This is especially clear in the social determinants of health literature, where the identification of pathways is seen by proponents and detractors alike as crucial for the case that the socioeconomic factors identified as health determinants really do have the effects claimed for them. Michael Marmot's Whitehall studies provide evidence for a causal link between social status and health, among British civil servants – a population whose basic biological needs (food, water, shelter) are amply met. Marmot's efforts to confirm this hypothesis have not focused merely on replicating the results in different populations (though that is of course one area of activity). A considerable amount of effort has also been devoted to identifying mechanisms by which social status might affect health.

This suggests a methodological thesis: that identifying an underlying mechanism is both necessary and sufficient for establishing a causal hypothesis.[4] I am not suggesting that Marmot or anyone else endorses this thesis: indeed it may be that no actual epidemiologist would sign up to it, in that blunt form. But identifying mechanisms clearly *is* relevant to establishing causal hypotheses in epidemiology, and setting up a somewhat extreme but clear stance may be a useful technique for exploring that relevance.[5]

It is not hard to see why identifying a mechanism might be considered necessary and sufficient for establishing a causal hypothesis, at least in epidemiology. An argument for its being necessary might appeal to the fact that epidemiology is clearly not a fundamental science. Causal associations identified in epidemiology presumably exist in virtue of the way that things are organized at a more fundemental level. The existence of causal associations at the population level is naturally seen as arising from certain regularities in the way that the members of that population are organised, and regularities in their

[4] Where there is more than one distinct mechanism underlying a given association, presumably all the underlying mechanisms would have to be identified to satisfy the spirit of this requirement.

[5] The mechanistic stance is *not* supposed to be the view that identifying mechanisms is the goal of epidemiological research: only of causal inference. Causal inference may have other goals, such as public health intervention, or indeed informing further causal inferences in an iterative process.

environment. Identifying a mechanism is just identifying the properties and activities of the population's members and environment which together give rise to the population-level association. If no mechanism can be identified, then the status of the population-level causal association remains mysterious. Failure to identify such properties and activities is not always evidence that they do not exist, of course, and perhaps this is why Austin Bradford Hill famously urges that 'biological plausibility' be treated with caution (Hill 1965). Nevertheless, a hypothesis for which no mechanism is *remotely* plausible, or for which no mechanism is discovered after a long period of time, remains at best tentative.

It is similarly obvious why identifying a mechanism underlying a causal association might be considered sufficient for establishing the corresponding causal hypothesis. If the events and activities giving rise to an association are identified, then it presumably follows that the variables which the hypothesis asserts are causally linked at least *can* be causally connected. There is an interesting twist here, however. Showing that a mechanism exists by which, say, stress can cause poor health, does not bear directly on the claim that stress *does* cause poor health in any particular population. The mechanisms identified by Marmot are ways in which the results of the Whitehall studies *might* have come about. This explains his two-part strategy outlined in the excerpt quoted previously, of first identifying mechanisms and then arguing for their actual operation in humans and animals. So identifying a mechanism is not on its own sufficient for establishing a causal hypothesis; a further inference is required to the claim that the identified mechanism is indeed the explanation of the causal association asserted by the hypothesis in question. What is sufficient, then, to establish a causal hypothesis, is the identification of the mechanism *actually* responsible for the association, not merely a mechanism which physically *could* be responsible for it. I take it that the fundamental motivation for this stance is that it follows from the claim that A causes B in a particular way, that A causes B simpliciter. Showing *how* A causes B is only possible if, or in other words entails that, A does in fact cause B.

The mechanistic stance responds as follows to the difficulties facing causal inference in epidemiology which I identified in Section 3.1. In answer to the question when we are justified in inferring causation for a studied group, the mechanistic answer is presumably, 'Not until a mechanism has been identified'. A single epidemiological study will not usually identify a mechanism, perhaps explaining why single studies typically provoke replication and further research, rather than a causal inference. But if epidemiology seeks to uncover mechanisms, then there may be another, more subtle reason that replication is important. Replicating in Study B an association observed in Study A provides evidence that the same mechanism underlies *both* associations; it seems prima facie less likely that two different mechanisms gave rise

to the two associations, than that one did.[6] This explains why replication by Study B can confirm a hypothesis about Study A: because Study B provides further reason to think that some mechanism underlies both studies, and thus provides further reason to think that some mechanism underlies the association first observed in Study A. And showing that some mechanism underlies an association is sufficient for showing that the association is indeed causal, on the mechanistic stance. This methodological stance therefore explains and vindicates the reluctance of epidemiologists to make causal inferences on the basis of individual studies, and the importance they attach to replication.

The mechanistic stance also provides useful clarification of the other difficulties we identified, the problem of generalizing from studies to population, and the question of how to interpret general causal hypotheses in epidemiology. Generalizing to a wider population is typically safer when the mechanism underlying a causal association has been identified, because knowledge of the mechanism yields detailed knowledge of what differences are relevant to the association. For example, our knowledge of the mechanism underlying the analgesic effect of paracetamol enables us to identify the circumstances relevant to this effect, and thus to say whether the effect will hold in a very wide range of circumstances. Moreover, it also gives us a lot of other useful knowledge, for example, about *other* associations – other effects of taking paracetamol. These uncontroversial facts motivate a methodological idea: that the generalisation of an association observed in a particular study to a wider population is only really warranted when the mechanism underlying the association has been identified. generalisation before a mechanism has been identified may of course be required in some circumstances, but until the mechanism has been identified, generalizing remains a sort of guessing, according to the mechanistic stance: because we can't be sure exactly *when* our causal generalisations will hold, until we know *why* they hold.

This suggests an interpretation of general causal claims, such as 'smoking causes lung cancer' or 'genetic influences on BMI and abdominal adiposity are high in children born since the onset of the pediatric obesity epidemic'. Such claims are to be interpreted, first, as asserting the existence of a mechanism – or perhaps several – linking the causal and effect variables, and second, as claiming that the operation of this mechanism(s) is (are) in fact what underlies the causal association between the two variables. The mechanism may or may not be known. When it is, this interpretation is very natural. 'Paracetamol causes pain relief' is naturally precisified, not by specifying more exactly the probability that an analgesic effect will be observed in various circumstances, but rather by saying more about the way paracetamol works. Detailed description of the mechanism *yields* information about whether an analgesic effect

[6] This is an application of what is sometimes called the Common Cause Principle.

will be observed in a wide range of circumstances, in an efficient way. When a mechanism is not known, the causal generalisation is a sort of stand-in: a claim that some unknown mechanism does link the variables.

The exception-ridden nature of epidemiological generalisations is unproblematic, on this view. Suppose a causal generalisation is just a claim that a mechanism underlies the association between two variables C and E. That is to say, in individual cases where an individual c leads to an individual e, a mechanism connects c and e. The exception-ridden nature of the generalisation reflects the fact that sometimes, C-events occur, but the mechanism by which they give rise to E-events is absent, and so E-events fail to occur. Likewise, the fact that C-events can by a certain mechanism give rise to E-events does not preclude E-events from coming about some other way. Causal generalisations are claims about the actual linking of individual C-events and E-events by actual instances of a mechanism; such claims are entirely compatible with C-events occurring without the E-producing mechanism, or E-events occurring as the result of some other mechanism. The content of the causal generalisation is that a certain mechanism *is responsible for the causal association between the identified variables*; it is not a claim about the association itself, but about how the association arises. Further support for this interpretation of exception-ridden causal generalisations derives from one of the points argued in the last paragraph, that when a mechanism is known, it is natural to precisify a causal generalisation by specifying the underlying mechanism, rather than offering more statistical detail about the circumstances in which the association holds. Whether or not it is a model for other sciences, it seems to be a good fit for epidemiology.

The one trouble previously identified which the mechanistic stance does not seem to handle is the worry about measures of proportion of influence, such as heritability. It can handle causal generalisations amounting to associations of less than 100%, such as 'smoking causes lung cancer', in the manner already indicated. But a claim such as 'pediatric obesity is over 70% heritable' does not submit to a mechanistic interpretation, since presumably mechanisms linking genes to weight exist in every human being. I am, however, inclined to think that this reflects badly on the concept of heritability, and similar attempts to apportion causal responsibility; such concepts are hard to make sense of in any analysis. It may be that they have no clear sense, or that they need further clarification (cf. Lewontin 1974; Schonemann 1997). Accordingly I propose to set them aside, and focus on interpreting those epidemiological hypotheses that clearly do make sense.

The mechanistic stance has a lot to recommend it, then: it provides a neat interpretation of several tricky features of actual epidemiological practice, and thus vindicates that practice. The idea that epidemiologists ought to first identify risk factors, and then look for the mechanisms underlying them, sounds like both sensible methodological advice and a fair description

of actual epidemiological methodology. In the next section, however, I will identify and amplify some doubts that some epidemiologists have expressed about the mechanistic stance.

3.3 Black boxes

Notwithstanding the foregoing, there are some reasons to doubt that the mechanistic stance provides a good methodology for causal inference in epidemiology. Both the necessity and the sufficiency of discovering a mechanism for inferring causation are open to criticisms of principle, and these criticisms can be illustrated by actual episodes in epidemiology.

Let us start with the idea that discovering underlying mechanisms is necessary for the inference of general causal hypotheses. The motivation identified in the previous section for this methodological principle is that the existence of a mechanism is necessary for the existence of a general causal relationship. On the mechanistic interpretation, a causal relationship exists at the population level between two variables only if the particular instances of those variables are related by a mechanism. I suggested that the general causal claim could be interpreted as a claim about the existence of an underlying mechanism. If so, it is natural enough to require that we *discover* this mechanism when we assert a general causal claim. So an inference to a general causal claim requires, as warrant, the discovery of an underlying mechanism – sooner or later.

This 'sooner or later' indicates a difficulty, though. Our failure to discover an underlying mechanism does not mean the mechanism doesn't exist. Moreover, when we are entirely ignorant of a mechanism, it seems that we are often quite incredulous about the existence of one. Combined with the view that a mechanism must underlie causal associations, this native incredulity can lead, and on occasion has led, to an unreasonable prejudice against hypotheses when we do not see how the hypothesised causal link might work. The view that mechanisms need to be identified for a causal inference to be secure can colour our assessment of the evidence for new hypotheses or against accepted hypotheses. I will suggest that this is because it is not really a methodological principle: it tells us what to aim at (discovery of mechanisms) but not how to achieve that goal. As a consequence, it tilts the balance in favour of existing knowledge, and inhibits what it recommends – the discovery of mechanisms about which we do not yet know.

Several well-known episodes from the history of epidemiology illustrate these claims. Perhaps most famously, the miasma theory of the nineteenth century offered a mechanism for the transmission of disease, based on the movements of 'miasms' (roughly, bad air). The fact that the theory purported to give a mechanism for disease transmission was its principle virtue.

Nineteenth century epidemiological heroes such as John Snow and Ignaaz Semmelweis were criticized for failing to identify plausible mechanisms for their causal hypotheses, leading to unnecessary loss of life in both cases. Snow argued, on the basis of incredibly careful door-to-door inquiries, that a causal connection existed between water supply and cholera (Cameron and Jones 1983). Semmelweis that differential childbed fever rates between two wards in a Vienna hospital were caused by the dirty hands of medical students, who worked in one ward but not the other (Carter 1994). Both hypotheses were resisted by the authorities and many doctors (although Snow was somewhat more persuasively successful than Semmelweis), and the principal reason given was that no plausible mechanism for the transmission of disease along these vectors had been identified. It was not until some decades later that a mechanism which might plausibly underlie their respective hypotheses was identified, in the shape of microbial theory (Carter 2003, esp. Ch. 3, 4).

J.P. Vandebroucke argues that Snow had a 'contagionist' hypothesis about the mechanism of disease transmission, and thus that this is not an example of 'black box' epidemiology (Vandenbroucke 1988). Nevertheless, it is doubtful that Snow had what we now view as a *correct* theory of cholera transmission (and contagionism is certainly false as a *general* theory, since it is possible to contract diseases other than from another diseased person, e.g. tetanus; puerperal fever). It is important to distinguish psychological claims about what leads scientists to their theories, from methodological claims about what justifies or otherwise warrants those theories. Snow's results are impressive *to modern epidemiologists,* even though we now believe he was wrong about the mechanism of disease transmission. His proposed mechanism cannot, therefore, be part of the warrant which we *now* accept for his conclusions (whatever he thought); yet we still regard those conclusions as warranted to a high degree by the evidence he procured. Semmelweis also hypothesised a disease mechanism, namely the resorption of animal-organic matter leading to the decaying of the blood (Gillies 2005). But similarly, acceptance of Semmelweis's resorption/decaying blood mechanism surely cannot be part of *our* reason for thinking that any of Semmelweis's causal hypotheses were well-founded, since we believe the resorption/decaying blood mechanism is not what underlay the associations he identified, for instance between disinfecting hands and reductions in differences in mortality between the two wards. Yet even though we reject his account of the underlying mechanism, this does not prevent us from accepting Semmelweis's causal hypothesis that disinfecting hands caused the reduction. Vandenbroucke's argument stumbles on exactly the point I am trying to make, confusing the discovery of mechanism as a *goal* of causal inference, with the discovery of mechanism as *method*. I am suggesting that it is a good goal, but a lousy method. It is not their theories of the mechanism of disease transmission which have elevated Snow and Semmelweis almost to hero status in the eyes of many

modern epidemiologists; and it is not the failure of their contemporaries to appreciate these mechanism-theories that is so often lamented. Rather, it is the way they procured evidence to support their causal hypotheses and to refute counter-hypotheses, which is so widely admired; and the way that evidence was ignored that is lamented.

Thus these episodes cast doubt on the usefulness of a methodological principle stating that discovery of mechanism is necessary for warranted causal inference. At least three doubts are distinguishable. First, the obvious logical point that has already been mentioned: that the inference from no known mechanism to no mechanism is a tricky one. Second, demanding that a mechanism be identified before a causal inference is accepted simply seems to be an unreasonable position, because it seems possible to have excellent evidence for a causal link, without understanding how the link works. Even if the mechanistic *interpretation* is plausible, and general causal claims are to be interpreted as existence claims about underlying mechanisms, it does not follow that a general causal claim is only warranted when the underlying mechanism is identified. It is possible, in an epidemiological context, to know that there is a causal link – and therefore that a mechanism must exist, on the mechanistic stance – yet not know what that mechanism is. The opponents of Snow and Semmelweis are generally considered to have been *unreasonable* to doubt the extremely convincing evidence for a causal link; this is difficult to explain if warrant for causal inference requires the identification of a mechanism.

Of course, in these cases, a mechanism *was* eventually discovered. But (and this is the third doubt) the discovery came *after* the causal hypothesis was well-established. The mechanistic methodology gets things the wrong way round, in these cases. The discovery of a mechanism *can* of course help to confirm a causal hypothesis, but a causal hypothesis can also be solidly confirmed well before the underlying mechanism is known. Therefore discovery of an underlying mechanism is not a necessary condition on warranted causal inference.

It must be remarked that drawing morals from historical episodes is a delicate business, because it is possible for different commentators to see different lessons in the same episode. For example, while most commentators (of a moral-drawing sort) would agree that something went wrong in the Semmelweis episode, Federica Russo and Jon Williamson see the episode differently. They insist that, to establish a causal claim, it *is* necessary to identify an underlying mechanism. But they do not adopt the mechanistic stance as I have outlined it, because they do not think that identifying a mechanism is sufficient for causal inference: they hold that it is also necessary to provide what they variously refer to as 'statistical' and 'probabilistic' evidence. In support, they cite episodes where causal hypotheses were supported by evidence of one kind but not of the other, and were rejected on that basis. Thus they cite the Semmelweis episode to illustrate the claim that 'the relation

between contamination and puerperal fever . . . was not accepted until backed up by mechanistic evidence, i.e. until the germ theory had been developed' (2007, p. 11), as an instance of their more general thesis that identifying mechanisms is necessary (though not sufficient) for causal inference. They go on to propose a theory of causation which is intended to fit this methodological picture.

This line of argument makes me uneasy, because I am unsure whether it is meant as a descriptive account of causal inference (then and now), or as a normative account of the standards which *ought* to be used when deciding whether to infer causation. Suppose we grant (for the argument) Russo–Williamson's descriptive claim, that Semmelweis's contemporaries rejected his theories because of a lack of 'mechanistic evidence', which I take to mean a lack of any then-acceptable theory about what the underlying mechanism for the proposed causal association might be. In this sense, indeed, Semmelweis did fail to establish his various causal hypotheses: he failed to provide evidence which was *in fact* compelling, as demonstrated by the fact that his evidence did not compel many of his contemporaries. But in another sense he clearly *succeeded* in establishing (at least some of) his causal hypotheses: he provided evidence which, in the eyes of most modern epidemiologists, his colleagues *ought* to have taken more seriously. In particular, the evidence for the efficacy of his proposed intervention – disinfecting (not merely washing) hands – is extremely strong. And replication would have made it stronger, without necessarily advancing knowledge of underlying mechanisms.

Unfortunately a purely descriptive reading of Russo and Williamson's claim renders it largely irrelevant from a methodological point of view, and does not justify or explain why they themselves treat it as a motivation to seek a theory of causation, apt for the health sciences. If the claim is merely descriptive, then we may conclude that Semmelweis's contemporaries were simply *wrong* to insist that a mechanism be identified before they accepted any causal connection. (And so we do not need a new theory of causation that would be compatible with this insistence.) Suppose, then, that Russo and Williamson intend their claim to be normative. Then they are making a normative claim that mechanistic (as well as statistical) evidence is necessary for *good, rational, warranted* causal inference. On this reading, the lesson Russo and Williamson draw from the Semmelweis episode is that Semmelweis's theories *ought* not to have been accepted until knowledge of underlying mechanisms was obtained. (This explains why they offer a theory of causation intended to justify this stance.)

But if this is indeed what Russo and Williamson are claiming, then I fear they need a far stronger argument than any they supply. For then they are committed to the startling view that, had germ theory not come along and the underlying mechanism remained a mystery, we today would be rational to dismiss Semmelweis's work, no matter how much evidence we had gathered

in the meantime about the efficacy of disinfecting hands. That is not a view which many modern epidemiologists would share. Modern epidemiologists set very high store in some methods, such as the randomized control trial, which involve no requirement to identify underlying mechanisms. Empirical evidence suggests that this view would cost lives if it were adopted by modern epidemiologists, and that indeed it did cost lives if it was in fact the reason that Semmelweis's claims were not accepted more promptly. In short, the descriptive Russo–Williamson thesis does not (without further argument) support any claims about causation, nor bear on questions about how causation *ought* to be inferred; while the normative thesis is false by the methodological standards prevalent in epidemiology, and also, arguably, in light of the empirical evidence from episodes such as the Semmelweis case itself.

In fact, on the mechanistic interpretation, it is not surprising that a causal hypothesis can be well confirmed before the underlying mechanism is discovered. On the mechanistic interpretation I have suggested, a causal generalisation asserts that there *is* some mechanism that is responsible for the association in question. We could have good reason to believe that there is some mechanism, yet not know what it is.[7] Moreover, this order of events suggests a plausible story about how we discover mechanisms when we previously had no idea about them. We make warranted inferences to causal generalisations; these generalisations imply the existence of underlying mechanisms; and we then conduct further research to find the mechanisms. We *know* where to look. It seems, then, that the mechanistic metaphysics does not after all motivate the corresponding methodological principle that the discovery of underlying mechanism is necessary for warranted causal inference.

Stepping for a moment beyond the confines of the methodology of epidemiology, there seems to be little intuitive support for the idea that causal inferences require knowledge of mechanisms as warrant. There are many everyday cases where we take ourselves to have knowledge of causal relations, without having the slightest idea about the mechanisms underlying them. I know that the clear Turkish liquor, rakı, goes cloudy when water is added, but I don't know what the underlying mechanism is. Presumably someone does, and it might be suggested that I can have a warranted causal belief just as long as I have recourse to an expert who can explain the mechanism to me. But the Turks have known for centuries that rakı goes cloudy when water is added. It is absurd to insist that, for centuries before the underlying mechanism was known, the Turks did not know that mixing water and rakı in roughly equal quantities caused the cloudiness they routinelys witnessed.

Moreover, to insist as a general matter that underlying mechanisms must be identified before a causal inference is warranted raises a dilemma. As we go

[7] This is an instance of the more general fact that we can know *that* something exists without knowing what it is. For example, I know that there *is* something holding the floor up, but I don't know what it is; you may know that there *is* something in the cave, but not what, etc.

on uncovering underlying mechanisms, either we will come up against causal relations for which no underlying mechanism can be discovered, or we will not. If we do, then none of our causal knowledge will be secured, because we will have reached a point where we are unable to discover underlying mechanisms and therefore unable to obtain the warrant we sought for our higher-level causal inferences.[8] If we do not, then we will never finish uncovering underlying mechanisms, and thus again we will never obtain the warrant we seek for our causal inferences. As a general methodological principle, then, the requirement that underlying mechanisms be identified, before a causal inference is warranted, is a guarantee that we can never have causal knowledge. If it has any applicability then it must be confined to particular domains, such as causal inference in epidemiology; but historical episodes previously alluded to suggest that it does not work even so confined.

These famous historical episodes have more recent echoes. In an influential (but not uncontroversial) report for the US government, Richard Doll and Richard Peto argued that many environmental causes of cancer could be identified from careful analysis of epidemiological evidence (Doll and Peto 1981). This approach suggests that epidemiologists should treat diseases as 'black boxes' (Peto 1984), and that the identification of a causal mechanism is not necessary for a warranted causal inference. Accordingly, epidemiologists need not concern themselves with the discovery of mechanisms, but can directly attack causal questions without worrying about the mechanisms underlying the hypotheses they generate. As in the cases of Snow and Semmelweis, the pragmatic benefits of this approach are evident. If Doll and Peto are correct, then labouring to uncover mechanisms may well prove to be a waste of time and money, from a public health point of view. Especially where environmental causes are concerned (smoking being the best-known example), refusing to make a causal inference until a mechanism is known can be seriously detrimental to public health. Doll and Peto's recommendation typifies what I will call the *black box stance*.

A great deal of contemporary research is, I think, undecided between the mechanistic and the black box stances. On the one hand, epidemiological research largely proceeds by identifying associations and applying various statistical tests and methodological principles to form a view about whether these associations are causal. On the other hand, the explosion of identified risk factors has not produced a corresponding increase in the scope of our

[8] Might it be objected that warrant of a higher-level causal hypothesis can be conferred by knowledge of an underlying mechanism, even if that underlying mechanism contains causal links for which yet lower-level underlying mechanisms are not known? No: because (on this view) we are not warranted in believing that the causal links in the mechanism underlying our higher-level hypothesis, until we have identified the mechanisms underlying them in turn. Without that warrant, we do not know (on this view) that they are causal links: for all we know (on this view), our putative mechanism may be a coincidental dance of its parts.

understanding of the conditions studied, nor has it been accompanied by a corresponding explosion in public health or medical interventions. This is not just a case of technology lagging: it is also due to the fact that the causal hypotheses in question do not seem terribly reliable. For example, studies seemed to show that hormone replacement therapy reduced risk of heart disease, and public health policies were implemented on this basis, before subsequent studies found the opposite effect (Rutter 2007). A slightly more subtle problem is that directly translating a causal hypothesis into a public health intervention may have unintended consequences (an instance of the generalisation problem). For example, beta blockers administered after surgery appeared to reduce the risk of death by heart attack, but subsequent studies showed that they increased the risk of death overall by increasing risk of death by stroke and other conditions (Devereaux *et al.* 2008). This, of course, is grist to the mill of the mechanistic stance, because one of the chief benefits of discovering the mechanism underlying an association is that it often comes with information about *other* associations, and so makes unintended consequences of this sort less likely.

Requiring that a mechanism be identified before a causal hypothesis is accepted may be too strict; but it does at least have the merit of clarity. It is easy to tell *whether* a mechanism has been postulated; moreover testing the hypothesis by replicating the *mechanism* may sometimes be a more straightforward (because lab-based or clinical) business than replicating the association itself in a large observational study. Whereas analysing the methodology of a published study in order to form a view as to the security of its results is devilishly difficult. Indeed it may be impossible, depending on the accuracy and completeness of the published methods section.

One solution to this tension, then, would be to require the discovery of mechanisms as an admittedly too strong necessary condition on causal inference. In pragmatic contexts where great harm appears to be a real possibility, but where a mechanism cannot be identified, some other decision-theoretic principle, such as the precautionary principle, might be appealed to. This stubborn mechanistic stance is, however, a last resort, because of the difficulties we have been discussing, and because identifying sound principles for decision-making under uncertainty is itself a difficult task. Even erring on the side of safety is not straightforward. The HRT/heart disease example shows that we can be wrong, not only in our causal inferences, but also in deciding which way it is safe to err.

So far I have been focusing on the view that mechanism discovery is necessary for causal inference. What about the claim that it is sufficient? Here, the mechanistic stance might appear to be on stronger ground: showing *how* A causes B seems to entail that A *does* cause B.

Unfortunately, that does not mean that 'look for mechanisms' is a good methodological principle. It has at least two serious weaknesses. First, it is

too vague. Ironically, it does not tell us *how* to look for mechanisms. It states a goal, but gives no indication how to get there. This, I suggest, explains why the mechanistic stance has on occasion led to a bias towards existing knowledge. Mechanisms we already take ourselves to know about satisfy the methodological directive; but the same directive doesn't help us find mechanisms we don't know about.

Second, the search for mechanisms can mislead, because it can allow us to believe we have achieved an understanding of something when we have not. Showing *how* A causes B indeed entails showing *that* A causes B. But giving information about an event's causes is not sufficient to explain that event, nor to allow you to devise an effective intervention. If I explain my late arrival by telling you that I was born, I am citing a cause of my late arrival: but I am probably not providing a good explanation of my late arrival. Similarly, identifying a mechanism by which, say, hydrochloric acid leads to ulceration of the stomach lining is indeed sufficient for showing *that* the presence of acid causes ulceration. But there is an explanation for the presence of excessive acid, in the case of many sufferers, namely the presence of bacteria, *Helicobacter pylori*. Showing that stomach acid causes ulceration by identifying the mechanism is a good method for proving causation; but epidemiologists (like other scientists) are typically interested in more than cataloguing the causes of the phenomena they study. They are interested in explaining them, and intervening to change them. Each of these goals plausibly requires the identification of causes. But not just *any* causes. The mechanistic methodology misleads because it provides a sufficient criterion for causation, but no guidance on whether the 'right' causes have been identified, or what the 'right' causes are.

This philosophical point is also well-illustrated by famous historical episodes. I have already mentioned the discovery of *H pylori*. Thoroughly documenting the mechanism by which hydrochloric acid causes stomach ulceration did not lead to the discovery of *H pylori*. Moreover, the hypothesis that peptic ulcer might in many cases be an infectious disease was initially treated with considerable scepticism – because it was thought that bacteria could not survive in such an acidic environment as the stomach. What led to the discovery of *H pylori* was initially chance observation of unknown bacteria in patients with peptic ulcer, followed up by observational studies, clinical work on the bacteria, and a dramatic piece of self-experimentation (for a summary see Angel 2008, Ch 2).[9] Of course, the discovery of the mechanism by which *H pylori* causes ulcer is also an important feature of this episode. I don't mean to deny that discovering mechanisms is important and useful: only that the directive to do so is not a reliable guide for causal inference in epidemiology. This is not because it does not establish causation, but because it does not identify the *right* causal associations. Identifying causes, *any* causes, is not

[9] One of the discoverers, Barry Marshall, drank a solution containing the bacteria and developed gastritis, then took antibiotics and recovered.

enough; there may be *other* causes – like *H pylori* – which better explain the phenomena in question, or offer readier foci for intervention. *Mere* causal inference is not all it is cracked up to be.

What makes the *H pylori* case such a neat illustration is that acid *does* play a role in ulceration. The bacteria cause excessive acid production, which is what directly causes ulceration. So there *is* a mechanism there. In the cases of Snow and Semmelweis, on the other hand, the mechanisms identified we would now regard as non-existent. So in those historical cases, a defender of the mechanistic stance might argue that, had the *real* mechanism of disease transmission been believed rather than miasma theory, the hypotheses of Snow and Semmelweis would have been better received. That may well be so. But this argument confuses reality with our grasp on it. A sound methodological principle cannot rely on our already knowing what we are trying to find out. In the case of miasma theory, many medical scientists *thought* they knew the mechanism of disease transmission. Partly as a consequence, they failed to properly appreciate the evidence before them. This is not simply a case of scientists being convinced of something and failing to give due weight to disconfirming evidence: it is a case of scientists believing they understand *how* something happens, and rejecting causal hypotheses that appear incompatible with this mechanism.

This problem seems to be at least partly what Doll and Peto have in mind when the advocate the black box stance. They argue that epidemiological evidence can warrant many causal inferences, without the underlying mechanisms being known. This sounds like a sort of call to arms for epidemiologists, a rallying cry for them to have faith in the methods of their discipline, and in particular to pay attention to the causal associations revealed by the evidence, without worrying about how the causal associations might work. Nevertheless, the black box stance does not offer much in the way of positive methodological recommendations. And it does not entirely dispense with the fundamental motivation of the mechanistic stance: that if we want to *really* explain – or control – something, we need to know how it works. In the next section I will propose a reconciliation.

3.4 Contrasts

It may be tempting to see the contrast between the mechanistic stance and the black box stance as a disagreement about the goals of epidemiology. One epidemiologist puts it like this:

> . . . the epidemiologist who tries to explain, and if possible eliminate, variations in disease occurrence without much regard for mechanisms, stands in contrast to the laboratory scientist who prefers to disentangle the mechanisms first.
>
> (Vandenbroucke 1988, p. 708)

The mechanistic stance might be seen as a more properly scientific view, interested in deep explanation; while the black box stance might be thought of as a more pragmatic view, interested primarily in designing public health interventions by the most direct inferential route available.

This is a misunderstanding, in my opinion. The strengths and weaknesses of the mechanistic stance apply equally to the goals of explanation and intervention. Let me enumerate the principle strengths and weaknesses of the mechanistic stance, starting with the strengths.

(i) Interpreting causal generalisations in epidemiology as existence claims about underlying mechanisms resolves gives a clear and plausible meaning to those generalisations, and helps us to understand the role of replication in epidemiology.

(ii) How widely we can generalize from a known association seems to be directly linked to how well the underlying mechanism is understood.

(iii) Discovering an underlying mechanism proves the truth of the causal hypothesis in question.

(iv) A mechanistic explanation of a causal association increases our understanding of that association.

(v) Knowledge of the mechanism underlying one causal association gives us, or at least can lead to, knowledge of *other* causal associations.

Each of these is a strength, whether the goal is scientific explanation or public health intervention. (i) is perhaps the most philosophical advantage, but it is surely of some importance to the scientist and the public health policy maker alike to have a good grasp on the nature of the causal generalisations they employ. For the scientist, it yields greater understanding; for the policy-maker, the ability to avoid practical consequences of misunderstanding. (ii) is evidently of central interest to the policy-maker, since how widely the results of a given study apply is central to the question of what interventions it warrants. It is also of evident interest for someone who is interested primarily in explanation, since it bears on how much a given causal hypothesis might explain. (iii) is of interest for both explanation and intervention, since false causal inferences are neither explanatory nor reliable guides for intervention. (iv) is evidently of value for those interested primarily in explanation, but it is also of use to the intervention-focused. This is because when we understand a mechanism, we are often able to identify more than one point at which we might intervene on that mechanism (if we want to prevent it) or more than one point at which it might be vulnerable to breaking down (if we want to protect it). (v) is useful from an explanatory perspective, since knowledge of other associations is a symptom of explanatory power, as well adding to the grand total of our knowledge. And it is evidently useful from the point of view of intervention, because it enables us to avoid unintended consequences.

A similar exercise shows the weaknesses of the mechanistic stance to apply regardless of whether explanation or intervention is the goal. To summarize the weaknesses:

(i) A general causal hypothesis can be warranted before the underlying mechanism is discovered (indeed, a warranted causal hypothesis is a great reason to look for a mechanism).

(ii) The directive to seek mechanisms is non-specific, and does not tell us how to find mechanisms; on some occasions, this appears to have led to a bias in favour of known mechanisms and against causal associations for which no mechanism is yet known.

(iii) A causal hypothesis may be deficient even though the underlying mechanism is understood, because it may lead us to believe that we have obtained a greater understanding of a phenomenon than we really have.

From an intervention-oriented point of view, (i) is a serious drawback of the mechanistic stance, because it shows that the mechanistic stance could lead to serious unnecessary delays on intervention. Less obviously, it is also a drawback from the explanatory point of view, because it leads to the unnecessary rejection of good explanations. From an explanatory point of view, the mechanistic stance is guilty of a why-regress fallacy. The 'why regress' is simply the fact that it is always possible to ask 'Why?', including on occasions when the explanation offered is a good one (Lipton 2004, pp. 21–2). What I call a why-regress fallacy is the refusal to accept an explanation on the grounds that the explanation itself has not been explained. As a general rule, such grounds are fallacious, because explanations can be good as far as they go, without providing the entire causal history of the explanandum. Otherwise, epidemiologists would study the Big Bang. (ii) is clearly a problem for any application of the mechanistic stance on causal inference, regardless of motivation. (iii) applies to intervention- and explanation-oriented foci equally, since it means that we miss out on potentially more fruitful interventions or explanations, respectively.

What, then, is the diagnosis of the tension between mechanistic and black box stances? I suggest it arises from a simple confusion of metaphysics and methodology. On the one hand, it does not follow from the strengths listed that discovering mechanisms is necessary for causal inference, nor that the discovery of mechanisms is a sufficient guide for explanation or intervention. None of these points directly motivate the mechanistic methodological stance on causal inference; what they motivate is *interpreting* causal hypotheses as existence claims about underlying mechanisms, and *seeking* these underlying mechanisms. Neither directive tells us anything about the *method* of causal inference; at most, they tell us about the goal. On the other hand, it does

not follow from the weaknesses listed that we should abandon a mechanistic interpretation of causal hypotheses, or that we should abandon the search for mechanisms. It only follows that we should not set the identification of a mechanism as a necessary condition for causal inference; nor confuse the power of a mechanistic explanation for a given causal association with the explanatory power of the association itself, with respect to the goals of our investigative activities.

The advantages of a mechanistic interpretation of causal hypotheses in epidemiology were laid out in Section 3.2. My suggestion is that we endorse this interpretation, but resist the temptation to take the mechanistic methodological stance with respect to causal inferences. The mechanistic metaphysics is good, but its methodology is bad.

The obvious next question is: how should causation be inferred? I do not think there is an easy answer. There are just a bundle of methods, organised around a common theme of identifying patterns of differences and similarities (Mill 1843; Lipton 2004), but increasingly statistically sophisticated (Spirtes, Glymour and Scheines 1993; Pearl 2000), and (it is to be hoped) increasingly reliable. The point of this paper is not to contribute a new method of causal inference, but to identify and debunk a tempting bad method. That method is what I called the mechanistic stance. But at the same time, I hope to have shown that a mechanistic metaphysics for causal generalisations has a great deal to offer epidemiology. There is, therefore, no need to *choose* between the identification of causal associations at the population level, and the identification of underlying mechanisms. On the mechanistic interpretation, causal hypotheses at the population level *are* existence claims about underlying mechanisms. There need be no opposition between epidemiologists conducting observational studies, and those trying to 'disentangle' mechanisms in a laboratory. They are studying the same thing.

It is not necessary to identify the underlying mechanism in order to have *warrant* for a causal hypothesis. But it is necessary in order to *explain* the association. A hypothesis can be perfectly warranted, without being understood. Conversely, a hypothesis may be well (mechanistically) understood, but may itself fail to provide a good explanation of the phenomenon in question for the purposes at hand. To prove a causal hypothesis, it is sufficient to identify an underlying mechanism. But identifying a mechanism is no guarantee of the explanatory power of the hypothesis itself – the explanatory power of the hypothesis that stomach acid causes ulcer, for example. Mechanism is sufficient for causation; but causation is not sufficient for explanation, or for purposes of intervention. Oxygen is a cause of every car crash, and I make that assertion not on the basis of any statistical information, but entirely on the basis of my knowledge of the mechanisms underlying internal combustion engines and human respiration. Yet oxygen offers good explanations of few,

if any, car crashes; and controlling the oxygen supply is not a particularly promising avenue for policy makers to pursue.[10]

What I have not done is offer any detailed analysis of the notion of a mechanism. The goal of this paper is to see what the notion can offer causal inference in epidemiology, not to analyse that notion itself. Nevertheless, I would like to finish by sketching a view of the relation between mechanistic explanation and other kinds of causal explanation. I am inclined to see mechanistic explanation as of a kind with causal explanation in general. I take it as widely accepted that causation is not sufficient for explanation, and that for a causal association to explain, it must amount to a difference between fact and foil (Lewis 1986; Lipton 2004, Ch 3). In a public health context, health provides a plausible contrast class, as I have argued elsewhere (Broadbent, 2009b). To give a mechanistic explanation of an association is to tell a story about how events of one type cause events of another, by filling in the intervening steps in the causal chain, and specifying what conditions must hold. It is tempting to see mechanistic explanations as causal explanations, with associations as their explananda. Presumably the relevant contrast class is the failure of that association to hold. Identifying possible failures is not purely a hypothetical exercise: typically there will be plenty of actual failures, since epidemiological generalisations are typically exception-ridden.

If this sketch of a contrastive analysis is approximately correct, it would explain why mechanistic explanations are so useful in epidemiology and public health: because they provide information about what makes the difference between cause-events leading to effect-events, and not doing so. And intervening on links of this sort is the central purpose of public health policy. However, this sketch also illustrates why mechanistic explanation is not necessary for good causal inference: because a causal difference between fact and foil can be identified, even before the causal link between the proposed explanans and the fact is understood. We can know *that* drinking dirty water is the cause of the difference between people with cholera and those without, even if we don't know how drinking dirty water results in cholera. And the contrastive approach shows why mechanistic explanation is not sufficient for effective intervention or explanation. If a causal hypothesis fails to identify a causal difference between good health and ill, then it will offer neither explanation

[10] Sometimes a distinction is drawn between distal and proximate causes of disease; and sometimes biological causes are also distinguished, which may be either distal or proximate. However, drawing and defending such distinctions is not easy, as a large literature in philosophy and also in jurisprudence shows (for references see, respectively: Broadbent 2008; 2009a). A discussion of these distinctions would not be relevant here, since there is no particular reason to hope that the most explanatory cause in a given circumstance will be either proximate, or distal, or biological. Moreover, even if there were some reason for epidemiologists to favour one of these categories, a choice would still have to be made among causes within them.

of nor intervention on that difference; and a mechanistic explanation of the causal hypothesis (e.g. miasma theory) cannot improve matters.

The correct analysis of mechanistic explanation is not, however, the purpose of this paper. To apply the moral reflexively, we can know that mechanistic explanation works, without knowing how it works. I hope to have shown the strengths of a mechanistic interpretation of causal hypotheses in epidemiology, and to have illuminated the pitfalls of a tempting but erroneous companion methodology.

Acknowledgements

I am grateful to Kevin Brosnan, Brendan Clarke, Donald Gillies, Stuart Glennan, Jeremy Howick, the audience at a conference on Causality and Mechanisms held in Kent 2009, and an anonymous referee, for their useful comments and criticisms.

References

Angel, Katherine (2008). Causality and psychosomatic histories in contemporary Anglo-American Biomedicine. Ph D thesis, University of Cambridge.

Broadbent, Alex (2008). The difference between cause and condition. *Proceedings of the Aristotelian Society* 108: 355–364.

Broadbent, Alex (2009a). Fact and law in the causal inquiry. *Legal Theory* 15: 173–181.

Broadbent, Alex (2009b). Causation and models of disease in epidemiology. *Studies in History and Philosophy of Biological and Biomedical Sciences* 40: 302–311.

Brunner, Eric, and Marmot, Michael (2006). Social organization, stress, and health. In *Social Determinants of Health*, ed. Michael Marmot and Richard Wilkinson, pp. 6–30. 2nd edn. Oxford: Oxford University Press.

Cameron, Donald and Jones, Ian G. (1983). John Snow, the Broad Street pump and modern epidemiology. *International Journal of Epidemiology* 12: 393–396.

Carter, K. Codell (1994). *Childbed Fever: A Scientific Biography of Ignaz Semmelweis*. Greenwood Press.

Carter, K. Codell (2003). *The Rise of Causal Concepts of Disease*. Aldershot: Ashgate.

Devereaux, P.J. *et al*. (2008). Effects of extended release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *The Lancet*. doi:10.1016/S0140-6736(08)60601-7. http://www.thelancet.com/.

Doll, Richard and Peto, Richard (1981). *The Causes of Cancer*. Oxford: Oxford University Press.

Farooqi, Sadaf, and O'Rahilly, Stephen (2006). Genetics of obesity in humans. *Endocrine Reviews* 27, no. 7: 710–718.

Gillies, Donald (2005). Hempelian and Kuhnian approaches in the philosophy of medicine: the Semmelweis case. *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 159–181.

Glennan, Stuart (2002). Rethinking Mechanistic Explanation. *Philosophy of Science* 69: S342–S353.

Hill, Austin Bradford (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* 58: 259–300.

Lewis, David (1973). Causation. *Journal of Philosophy* 70: 556–567.

Lewis, David (1986). Causal explanation. In *Philosophical Papers, Volume II*: 214–241. Oxford: Oxford University Press.

Lewontin, Richard (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics* 26: 400–411.

Lipton, Peter (2004). *Inference to the Best Explanation*. 2nd edn. London: Routledge.

Marmot, Michael (2006). Health in an unequal world: Social circumstances, biology, and disease. *Clinical Medicine* 6, no. 6: 559–572.

Mill, John Stuart (1843). *A System of Logic, Ratiocinative and Inductive*. 8th edn. New York: Longman's, Green, and Co.

Pearl, Judea (2000). *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Peto, Richard (1984). The need for ignorance in oncer research. In *The Encyclopedia of Medical Ignorance*, ed. R. Duncan and M. Weston-Sraim, pp. 129–133. Oxford: Perganon Press.

Rothman, Kenneth J., and Greenland, Sander (2005). Causation and causal inference in epidemiology. *American Journal of Public Health (Supplement 1)* 95, no. S1: S144–S150.

Russo, Federica, and Williamson, Jon (2007). Interpreting causality in the health sciences. *International Journal of the Philosophy of Science* 21, no. 2: 157–170.

Rutter, Michael (2007). *Identifying the environmental causes of disease: how should we decide what to believe and when to take action?* The Academy of Medical Sciences.

Schonemann, Peter H. (1997). On models and muddles of heritability. *Genetica* 99: 97–108.

Sesardic, Neven (2005). *Making Sense of Heritability*. Cambridge: Cambridge University Press.

Spirtes, Peter, Glymour, Clark, and Scheines, Richard (1993). *Causation, Prediction, and Search*. New York: Springer-Verlag.

Vandenbroucke, Jan P. (1988). The 'causes of cancer': A miasma theory? *International Journal of Epidemiology* 17, no. 4: 708–709.

Wardle, Jane, Carnell, Susan, Howarth, Claire M.A. and Plomin, Robert (2008). Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *American Journal of Clinical Nutrition* 87: 398–404.

·4·
Causal modelling, mechanism, and probability in epidemiology

Harold Kincaid

Abstract

This chapter looks at interrelated issues concerning causality, mechanisms, and probability with a focus on epidemiology. I argue there is a tendency in epidemiology, one found in other observational sciences I believe, to try to make formal, abstract inference rules do more work than they can. The demand for mechanisms reflects this tendency, because in the abstract it is ambiguous in multiple ways. Using the Pearl directed acyclic framework (DAG), I show how mechanisms in epidemiology can be unnecessary and how they can be either helpful or essential, depending on whether causal relations or causal effect sizes are being examined. Recent work in epidemiology is finding that traditional stratification analysis can be improved by providing explicit DAGs. However, they are not helpful for dealing with moderating variables and other types of complex causality which can be important epidemiology.

Introduction

This chapter focuses on the three subthemes of this volume – causality, mechanism, and probability – largely through the lens of recent causal modelling approaches in epidemiology combined with some general morals from the philosophy of science. The general morals concern tendencies in the sciences to try to make formal methods do more than they can and to down play domain-specific substantive assumptions in scientific inference, a process sometimes called 'black boxing'. The issues about mechanisms that I pursue largely concern claims by epidemiologists (Hafeman & Schwartz, 2009), claims echoed by social scientists (Hedström & Swedberg, 1998; Morgan & Winship, 2007), that various explanations in their fields are inadequate because they lack mechanisms. I use causal modelling results and my philosophy of science morals to help evaluate that criticism and to show some ways in which traditional epidemiology – analytic stratification analysis without use of explicit causal models – sometimes tries to get more out formal inference methods than they can really yield and how some recent uses of causal

modelling in epidemiology does the same. One particular instance of this over extension, I argue, comes from appealing to probabilities in epidemiology when they are ungrounded.

Section 4.1 explains the framework and some general morals about mechanism, causation and probability. Section 4.2 looks at the standard epidemiological practice of identifying risk factors by stratification analysis. I argue that traditional epidemiology tries to get by with very little in the way of causality and makes much less use of probability than it advertises. Section 4.3 turns to more recent developments in epidemiology employing Pearle's graphical approach to causality. I discuss both its strengths and limits in understanding the place of mechanism and probability in analysing causality in epidemiology.

4.1 A framework

Some philosophical preliminaries will be useful for framing the kind of issues I want to discuss in epidemiology. There are two truths from the past decades of history and philosophy of science about scientific method and explanation that I would defend: (1) that scientific inference and explanation cannot be adequately captured by an a priori, domain independent logic of science or, put in a positive vein, domain specific, substantive assumptions play key roles in scientific inference; and (2) scientists often believe or act as if they believe that their results are the product of precisely such a logic when they are not.

There are many lines of reasoning to (1). They include:

(a) The argument ad Carnapium given by Quine: if Carnap could not find a successful inductive logic, then there is not one or, more seriously, informative general inductive logics have not been forthcoming.[1] The causal modelling techniques discussed later in the chapter illustrate the situation: there are useful things that can be said using them, but their range of application is restricted because of the strong assumptions and prior background knowledge needed.

(b) The argument from holism: every part of the web of belief is at least indirectly connected to every other, and given a big enough empirical change in one part of the web, we might have to give up those parts of the web that look like logical truths. Thus no inference rule is indefeasible (Quine & Ullian, 1970).

[1] I take it that 'do not violate the probability calculus, first order logic, etc'. is not informative in that it does not say anything about standard methodological disputes, for example on the importance of novel data. I have argued elsewhere that Bayes' theorem is uninformative in this way (Kincaid, 2002).

(c) The argument from underdetermination: there are always possible alternative theories compatible with any given set of data and therefore there cannot be a logic of inference that tells us which is correct, given those data (Longino, 1990).

(d) The argument from conceptual humility: There is no reason to think our concepts, including the concepts of justification, rationality, explanation and other epistemic concepts, have a logic of necessary and sufficient conditions that determine their domain of application (Wilson, 2006).

(e) The argument from history and social studies of science, which is really many related arguments or pieces of evidence about science in practice: Specific instances of inference rules such as parsimony or inference to the best explanation turn out to involve non-logical, domain substantive empirical assumptions (Day & Kincaid, 1994; Sober, 1988). Scientific experiments in our best science seem to be declared decisive by a complex social process that looks contingent (Gallison, 1987). Apparently general scientific inference rules or virtues can conflict and have to be balanced on substantive grounds (Kuhn, 1977). Scientific papers do not reflect the process and uncertainties of the reasoning that went into them. There can be science that is well done by the most obvious standards such as replication, peer review, etc. and yet that involves non-epistemic values (Longino, 1990).

Obviously these arguments are of varying quality, clarity, and upshot. In particular, it is important to note the difference between conclusions about what cannot in principle be done anywhere in science and conclusions about what some specific research programme currently cannot do in practice. Science is a multifaceted process, and pronouncements about it as a whole in fact violate the pragmatist spirit that motivates many of the arguments given above. Of course, rules of inference that approach a logical status are great when you can get them, because if usable by real humans, they guarantee reliability. However, the above arguments are at least some reason to think that we should be alert to the substantive, domain-specific knowledge that is often needed for successful inference.

It is when conclusions are claimed to follow from rules alone, when in truth they do not, that problems can begin. My claim (2) above that scientists make such exaggerated assertions is an empirical one. One line of evidence that will be further supported in this paper comes from the widespread abuse of statistical significance in the social, behavioural and biomedical sciences. Other evidence is found in the studies cited in (e) above. On common-sense grounds it seems clear that scientists have to use methods developed by others without fully understanding how they work because of the scientific division labour. Not understanding the full details of a method can lead to

underestimating its limitations. It is the hope for and the difficulties of finding tight inference rules that I shall use in framing my discussion of mechanism, probability, and causation in epidemiology.

Some general morals follow from the above perspective that I will argue are equally applicable to epidemiology. The first is the now perhaps standard truism of Cartwright's (Cartwright, 1994) that if there are no causes in, there will be no causes out. There is no logic that gets you from probabilities or associations alone to causal conclusions. However, at this level of abstraction, the claim is not that interesting (if the premises do not use the predicate 'cause' it is not going to be in a validly drawn conclusion). The interesting question is which causes in are needed to get what causes out. That is an important, non-obvious question of real import to epidemiology and one that will be discussed below.

Another moral concerns the place of mechanisms. Asking that mechanisms be provided is a demand that can mean multiple things resulting in different claims:

1. In the philosophy of science literature (Bechtel, 2008), mechanisms are usually thought of as the component processes realizing some higher level capacities, e.g. the mechanism of memory. I call these 'vertical' mechanisms'. However, many requests for mechanisms are about providing intervening or mediating variables between a putative cause and its effect. Call these 'horizontal' mechanisms. The latter is generally more relevant to epidemiology, particularly to the causal modelling work I shall discuss.
2. We can also want mechanisms for explanation as opposed to confirmation. Those who want mechanisms because they might rule out spurious causation are targeting their role in confirmation and are generally invoking horizontal mechanisms. These are the kinds of concerns that I discuss in epidemiology and causal modelling. The philosophy of science literature on mechanisms largely emphasizes the idea that mechanisms are needed to provide sufficiently deep explanations, working with at least implicitly an ideal of the 'full' explanation. The work in epidemiology that I discuss does not generally come from this motivation.
3. Mechanisms might be important for establishing that there exist causal relations between variables or for establishing causal effect sizes. Their 'importance' might mean knowing the mechanisms is essential or not essential but useful.

Identifying these distinctions helps support my general doubts that there are useful universal methodological rules when it comes to the demand for mechanisms. There are just too many free parameters in that demand for us to take it as binding in the abstract. So, for example, consider the claim that

identifying mechanisms is necessary for ruling out spurious causation. If I am worried that the association observed between A and B might not be evidence for causation because they might be the common causes of some third factor, it may not help me to know how A and B's properties are realized, i.e. to know the vertical mechanism. What is needed is not the microdetail but evidence about possible other causes at the level of A and B.

I next want to draw some limited morals about probability from the general framework. We certainly do have clear formal rules for handling probability, and those are universal and a priori if anything is. However, their range of applicability may not be, and we must again be aware of extension beyond reasonable bounds. Arguably the two most basic situations grounding probability judgements come from either (a) there are chance mechanisms, processes, or devices or (b) measurement of consistent degrees of belief. Random sampling from a population, random assignment to treatment or control groups, or random measurement errors are the main instances of the first category and Bayesian updating of the second. We have to assess case by case which of these groundings, if any, is applicable. This may all seem obvious, but we shall see in that epidemiology sometimes uses probability notions where none of these groundings exist.

4.2 Traditional epidemiology

Much of standard epidemiology – what I will call 'traditional epidemiology' in contrast to newer work with explicit causal modelling that I discuss later – consists in identifying risk factors in observational data for disease via stratification analysis to eliminate confounders. I want to sketch some typical work in this vein to serve as the source of my general discussion. My focus is on the epidemiology of leukemia and, more specifically, the roles of benzene exposure and diet in leukemia.

Benzene is a hydrocarbon extracted from petroleum for a variety of uses. The first cases reporting a link between benzene and leukemia date back to the 1920s. It was only in the 1960s that evidence beyond case reports began to appear. Supporting studies are either case-control studies or cohort studies (Glass *et al.* 2006). Case-control studies identify a set of present or past cases of the leukemia and compare benzene exposure in that group to exposure and disese in a control group. Cohort studies follow a group of individuals over time, tracking exposure and disease status. A standard outcome measure is the odds ratio, which is the odds of disease in the exposed divided by the odds of disease in the non-exposed. The data are typically analysed in two ways. In simple stratification analysis correlation coeffecients are calculated within the relevant stratifying subset, for example potential exposures to other carniogens, producing a new estimate of the odds ratio. More complex analysis make

use of multiple regression, where other possible 'risk factors' are included and the adjusted odds ratio is reported along with either a signficance level or confidence interval for the association. Benzene, for example, is consistently statistically signficantly associated with chronic lympahtic leukemia but the associations with other leukemias do not usually meet standard significance levels and are described as 'not significantly associated with' these diseases.

Numerous studies have also 'implicated' (a standard wording) diet in cancer. The connection between diet and leukemia, however, is based largely on the one study of Jensen *et al.* (2004). That study identified cases of childhood leukemia in a northern California registry and interviewed all who agreed to participate (83%) about diet of the mother during pregnancy. Controls were selected from the same geographic area based on birth certificates and were matched on variables such as race. Again odds ratios were caculated, multiple regressions run with other risk factors, and associations reported with significance levels. Consumption of fruits and vegetables was inversely associated with acute lymphatic leukemia, the more agressive form of lymphatic leukemia commonly found in children. Benzene exposure was not ascertained.

These kinds of studies are a dominant form of inquiry and research reporting in epidemiology. They predominate in the journals. The main textbooks (Rothman, Greenland & Lash, 2008) consist mostly of discussion of the techniques for doing these kinds of studies. Their general form – (a) reporting multiple regression correlations from (b) observational samples relying on (c) tests of significance and confidence intervals as the acceptance criteria – is also common across the social sciences. Probability, mechanism, and causality get relative little focus in this traditional approach, and the discussions they do get are skewed by the logic of science gloss I mentioned above, or so I want to argue.

A somewhat curious element of this traditional practice is that it is acausal. The Glass *et al.* study is illustrative. The word 'cause' is used only twice, both times in the initial background discussion referring to the work of others. The paper's conclusion is repeatedly stated as establishing an association between low levels of benzene and leukemia. The standard risk factor analysis paper that makes up the majority of the published work in epidemiology shares the same trait: results are always reported as associations, not as causes.

Is this eschewal of causal conclusions merely reasonable humility about the limits of inferring causes from correlations? No doubt it in part is.[2]

[2] A referee wondered whether the reluctance to use causal language was counter-evidence to my thesis that there is a tendency in epidemiology, as in all science, to extend formal methods beyond their reasonable reach. In so far as epidemiology really is motivated by that goal it is indeed and that is a good thing. However, as I point out, despite the admonition to not confuse correlation with causation, epidemiology in the end makes lots of causal claims based on associations. What is exciting about work in causal modelling is that makes such assertions possible in a disciplined way, but also shows when those implicit moves are illegitimate.

However, the roots go considerably deeper than that. Historically, the origins of epidemiology are closely tied to 'positivist' doubts about causation as a legitimate scientific notion. Karl Pearson was extremely influential among the early practitioners of epidemiology. In his *Grammar of Science*, Pearson (1900) argued that the concept of cause had no place in modern science. Causation was too metaphysical a notion; association, on the other hand, could be given full mathematical rigour and should replace causal talk in modern science. So from the start epidemiology was built on correlations taken as ends in themselves.

Moreover, the prejudice against causes is built into the analytic methodology of epidemiology. The ubiquitous stratification techniques actually can be inconsistent with a causal interpretation. Standard practice is to report relative risks or odd ratios after adjustment for any factor that might be thought to change the size of relevant risk. Confounders are often defined in purely statistical terms – confounding of an association between two variables occurs when there is a third factor or variable that, when controlled for, changes the value of the correlation between the two variables at issue. If one were looking for causes, such a procedure would be guaranteed sometimes to produce wrong results. Multiple causal interpretations are possible, as I will discuss in more detail below, when controlling for a third variable changes the correlation between two variables under study. Conditioning on the common effect of two independent variables creates correlations that do not represent causal influence. Conditioning on a moderator variable – one that influences the effect size of another cause (more on this later) – reduces correlation as does conditioning on a common cause or an intermediate cause. So the very procedures that are used by themselves have no consistent causal interpretation. No doubt some or many epidemiologists realize this on some level, and I will report on their attempts to get clear on a better notion of confounding in the next section.

Thus this attempt to stick to associations alone is really not sustainable, for both pragmatic and theoretical reasons. The pragmatic reasons come because epidemiology wants to be relevant to policy – whether it is governments intervening or individuals deciding how to behave – and interventions concern what can be causally influenced. So a typical epidemiological report will describe only associations and provide no explicit causal model, but then conclude with something relevant to policy. The theoretical reason that epidemiology cannot be consistently associationist and non-causal is that associations alone are unacceptably arbitrary. Associations are always associations in a population or sample from a population. If the population is not in some sense a causally homogeneous one, then indefinitely many uninteresting associations of the 'coffee users on Tuesday have less leukemia' sort can be found. The number of associations is restricted only by our ability to imagine possible predicates or categories.

I would use this last point along with several others to argue that the place of probability in actual epidemiological practice in more minimal than might be imaged. My worry about arbitrarily many associations could be put as a variant of Cartwright's dictum 'no causes in, no causes out' that reads 'sometimes, no causes in, no objective probabilities out'. An objective probability, I take it, is one that picks up a real distinction in nature. Cashing out 'real distinction' is of course a matter of controversy, but grounding inferences to other populations seems essential to the notion. Doing that requires us to think there is a causal process behind the probabilities or correlations that we identify. That does not mean that objective probabilities or what philosophers would call 'nomic' generalizations or correlations must always represent a cause – generalizable correlations can result from a causal process involving colliders, for example, which will generate objective non-causal probabilities if the same causal process can be found out of sample. However, I believe that the associations of epidemiology, stripped of any causal basis, may not provide for reliable inferences to populations beyond those where the associations are initially found, because there are a great many accidental associations in any population. Probability talk in such circumstances is misleading, which is precisely a standard critique of probabilistic accounts of causation.

I also am suspicious that the main targets of epidemiological explanation – generic risk claims – are probabilities, though I admit that they may seem to be (see Russo & Williamson, 2007). The kinds of things epidemiologists want to explain are relative risks, odds ratios, and population attributable fractions of disease. These are not measures that vary from zero to one. They are based on frequencies, for sure, and can be converted to percentages. However, though they are loath to say it until that final paragraph with the policy and behaviour implications, what epidemiologists really want these to measure is effect size, a causal notion that need not be cashed out in probability terms. Relative risk can make sense in a single fixed population with deterministic causes that results from no sampling distribution or random assignment. Use of epidemiological information to make risk claims about individuals in the process of diagnosis may well be probability claims, but they are part of clinical medicine, not epidemiology proper.

Not only can measures of relative risk make sense in such populations, these kinds of populations are predominantly what epidemiology deals with. This fact provides another reason to think that probability plays a more limited role in epidemiology than it might seem. Most epidemiological studies do not involve samples picked randomly from a population (and generally randomized experiments are thought not to be part of epidemiology with the exception of clinical epidemiology). Take the work on leukemia cited above. None of it involves random samples except on a few occasions for the control groups, and then the samples come from a 'population' that is in effect a convenience

sample. This work shows that diet has some connection to leukemia in those individuals living in northern California willing to participate in a study. No one would pretend these are random samples from anything.

Epidemiologists nonetheless report significance levels taken as the probability of seeing a given result when there is no real correlation. However, we have to ask 'no real correlation where?' Since these probabilities are not explicitly taken to be subjective degrees of belief, their interpretation remains unclear. The closest I can find to a coherent answer in this regard is that the population is some hypothetical super-population from which the population under study is an instantiation and random sample (Morgan & Winship, 2007). However, that framework is to my knowledge notably understudied. Why, for instance do we think that the population is a random sample from the hypothetical population? And why do we want to make inferences about hypothetical populations anyway, unless we are doing so to talk about counterfactual causal possibilities rather than sampling error?

However, there is a natural Bayesian interpretation of the probability claims made by significance tests for these nonrandom samples. I can reasonably ask what probability I should attach to the claim 'if there were some randomizing process such as measurement error that was involved in generating my data, then it is probable/improbable that I should see data like this'. One can then use objective mathematical facts about the hypothetical source or error to assign a conditional probability, which is objective not in the sense of having been generated by a real mechanism but in the sense of following deductively from assumptions.[3] However, traditional analytic epidemiology is decidedly non-Bayesian.

That non-Bayesian commitment is also illustrated by the other ways that probability is minimized in epidemiology, namely, in the studied avoidance of using Bayes' theorem to make sense of results. Like much other biomedical and social research, epidemiologists sometimes use significance levels as straight indicators of probable truth or falsity. This, of course, contravenes Bayes' theorem in that both the prior probability of the hypothesis and the likelihood of the data on the maintained hypothesis – the power – are needed to interpret a significance level. Note that the role attributed to significance levels does not result simply from a reversion to subjective priors, for power calculations are rare as well. Rothman *et al.* (2008) devote three pages to the concept.

These familiar practices seem to me a clear instance of hoping that significance testing can provide a logic of inference without the need for further

[3] A standard way to do this, one not often invoked in epidemiology, is through what is called permutation analysis. For example, suppose one has values for treatment for cases and controls. The mean of observed differences is then compared to the distribution of the mean that is produced if the labels case and control are switched in many different ways.

substantive knowledge. Standard practice portrays itself as rule driven inference when it clearly cannot be, at least where the rules are valid.

Pushing the formulas to do more than they can also shows up when what I have been calling traditional epidemiology does try to talk about causality. One practice that epidemiology shares with the social sciences and other biomedical fields is the use of the R-squared statistic as a measure of how well the given causes account for an outcome. A second practice more specific to epidemiology is reporting what is called the 'population attributable fraction' as a way of measuring how much of the disease burden is due to some risk factor. Both are calculated with the correlational evidence that is common to traditional epidemiology. Not surprising, as purely statistical measures, neither is a reliable guide to causal importance.

Put in the usual regression terms, *R*-squared – 'the explained variance' – is calculated in terms of the predictive errors of the regression. It is the squared ratio of the covariance to the product of the standard deviations:

$$R^2 = \frac{\text{Cov}(X, Y)}{\text{StdDev}(X) \times \text{StdDev}(Y)}.$$

It is common to take a high *R*-squared to mean that the causal factors included in the model capture most of the causal influence.

However, this causal interpretation is more than the formula can warrant. The formula gives a statistical measure of how close the data points are to the regression line estimated from them. In the simple case where X causes Y with no other causes involved and we regress percentage changes of Y on changes in X with measurement error, it is obviously the slope of the regression line – the percentage change in Y that is associated with a percent change in X – that measures the size of the causal effect. The data points may be close to a regression line with a shallow slope and they may be far from one with a steep slope. The *R*-squared statistic is orthogonal to measures of causal influence.

It is not that this confusion has gone unnoticed. It has been. Rather, my point is: the hope for purely formal criteria leads to using formal measures beyond their legitimate domain of application. Warnings about interpretation are ignored.

The 'population attributable fraction' is a statistical measure specific to epidemiology that compares the amount of disease burden in a population exposed to a risk factor to the burden with no exposure. Thus it is:

> $(\text{P}(D)\text{P}(D|E-))/\text{P}(D)$, where P(D) is the (unconditional) probability of disease over a specified time period, and $\text{P}(D|E-)$ is the probability of disease over the same time period conditional on non-exposed status.

There has recently been a large debate between obesity researchers and researchers on other major causes of diseases such as cancer and

cardiovascular disease over whose disease contributes most to mortality (Flegal *et al.* 2005). The measure being used in the debate is the population attributable factor.

The population attributable fraction is no better a measure of causal importance than *R*-squared (Levine, 2008). People are subject to overlapping risk factors for the same condition. If we assess them one by one for causal importance in the manner recommended by this formula, the total causal contributions will sum to more than one or, put alternatively, there will be more explained deaths than there actually are.

Let me now turn to the role of mechanisms in traditional epidemiology. Needless to say, if traditional epidemiology avoids causal claims, then it is likely to avoid mechanisms as well. So it does. Here, however, I would argue that it is on better grounds than with its approaches to probability and causation, and indeed for causal reasons that epidemiologists invoke indirectly.

Consider the cases of benzene, diet, and leukemia cited above. I think they nicely illustrate my point (Ok, they were selected to do so) that whether mechanisms are needed for respectable science depends on the context. Currently there is mostly only speculation about the molecular route by which benzene might cause cancer (Atkinson, 2009). Actually the problem is not that routes are hard to picture but that there are too many imagined routes and not much evidence to pick among them. The molecular changes involved in cancer are enormously complex and diverse, so much so that there are reasons to doubt that the cancers form much of a natural kind (Kincaid, 2008). There are many routes to tumorigenic transformations and various metabolites of benzene could be involved in various of those routes. There is no definitive evidence for any of these possibilities as the mechanism by which benzene causes cancer, and most of these possible pathways have only been understood in the last decade.

Yet arguably the evidence was good that benzene causes cancer some time ago. Although the traditional epidemiological reports shy away from causal claims, they do provide evidence that allows for causal interpretation. The cohort and case control studies can mimic, if not ideally realize, the logic of the clinical trial. They can do so by showing that differences in exposure are associated with differences in outcome and then arguing that there is no third factor that might explain the association. The clinical trial logic is precisely designed to establish causality without having to understand the intervening steps or mechanism. Of course, the evidence is fallible and adding in the mechanisms to the story would strengthen the evidence in various ways we will discuss in the next section.

The dietary case tells a different story. The mechanism is not known and there are not many concrete ideas about how diet would influence leukemia. One concrete hypothesis is that foods with DNA topoisomerase II (DNAt2)

inhibitor eaten during pregnancy reduce DNA damage. However, in the study described above, when the subset of possible protective dietary factors was restricted to those with DNA topoisomerase II (DNAt2) inhibitor, the inverse correlation with leukemia lost statistical significance. We have considerably less confidence in the plausibility of a mechanism in the case of diet than in the case of benzene.

This weakness in the evidence takes on considerable importance because the correlational evidence for a link between diet and leukemia is much shakier than in the case of benzene. The number of studies is small and the association between diet and leukemia is not always seen. The effect is likely to be small compared to that of benzene and thus proportionately harder to find. The number of possible confounding variables is large and it is hard to make a case that they have all been controlled for. As we noted, the study described here did not control for benzene exposure, certainly a possible confounder. Diet is in much greater need of a mechanism if we are going to label it as a cause of leukemia.

4.3 Causal modelling

I turn in this section to explicit causal modelling efforts in epidemiology, with the focus again on themes related to mechanisms, causality, and probability. I ask to what extent the Pearl (2000) programme can shed light on the need for mechanisms, on the role of probability in identifying causality, and the limitations of the Pearl approach for some epidemiological questions.

A central and intuitive element of the Pearl programme is the directed, acyclic graph (DAG), an instance of which is given in Figure 4.1. The graph is directed in that there are arrows between the variables representing causal relations. A cause is direct if does not go through another variable or 'node' to

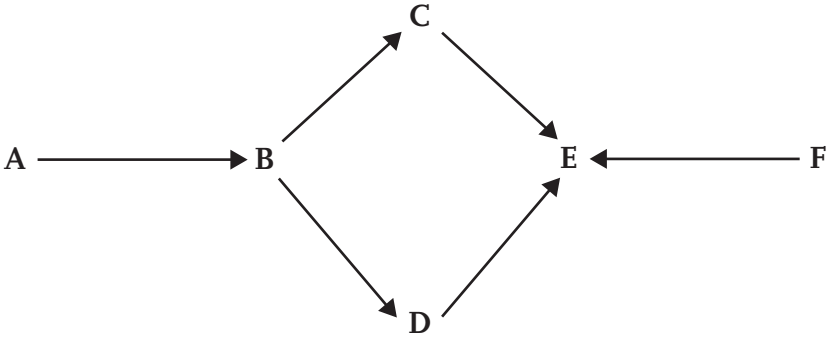

Fig. 4.1 Directed acyclic graph (DAG) with six variables.