Proof and Uncertainty in Causal Claims

Martine Jayne Barons* and Rachel Lynne Wilkerson

Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK.
*Correspondence: martine.barons@warwick.ac.uk

Abstract *Causal questions drive scientific enquiry. From Hume to Granger, and Rubin to Pearl the history of science is full of examples of scientists testing new theories in an effort to uncover causal mechanisms. The difficulty of drawing causal conclusions from observational data has prompted developments in new methodologies, most notably in the area of graphical models. We explore the relationship between existing theories about causal mechanisms in a social science domain, new mathematical and statistical modelling methods, the role of mathematical proof and the importance of accounting for uncertainty. We show that, while the mathematical sciences rely on their modelling assumptions, dialogue with the social sciences calls for continual extension of these models. We show how changing model assumptions lead to innovative causal structures and more nuanced casual explanations. We review differing techniques for determining cause in different disciplines using causal theories from psychology, medicine, and economics.*

Keywords: causality; Bradford Hill criteria; instrumental variables; causal algebras; graphical models; Bayesian Networks; mental models; uncertainty; proof.

Introduction

When can we say that one thing is the cause of another? In common parlance, we usually intend to convey that cause is a necessary and sufficient precursor of an effect. While this question motivates much of scientific research, causality may not be immediately amenable to the rigour and certainty of mathematical proof. Rather, some probabilistic, interventionist notion of causality is required to model the observations and uncertainty in the system.

In the search for truth and evidence, typically, we want to understand the real world and so we build a model that captures the entities, stimuli, relationships, and behaviours we observe (**French, 2015**). This can be a simulation or a mathematical model that describes the observed elements of interest. However, uncertainties arise because the model can only ever be an approximate representation of the world. There may be uncertainty about scientific theory, the strength of the effect, randomness, unknown

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

72

future values, and calculation accuracy. Additionally, all models incorporate subjective uncertainties, degrees of belief and preferences automatically built in by the choices and assumptions of the modellers. This leads to uncertainty about the descriptive model's ability to capture all the salient features of the world, uncertainty pertaining to the beliefs and values encoded within the model and uncertainty about how many analyses to perform to be sure of the model.

The lack of uncertainty in mathematical proofs, and their enduring nature is what attracts some to the field (**Barons and Chleboun, 2015**). Woodward, in his account of interventionist causation states that, 'genuinely explanatory proofs are those that show us how the truth of some theorem depends on the assumptions from which the theorem is proved.' Thus, 'when a theory tells us how Y would change under interventions on X, we have (or have material for constructing) a causal explanation' (**Woodward, 2003**). Proofs that are truly explanatory characterise a property about a structure in a theorem such that it is evident that the result depends on the property. Proofs of this nature allow us to see how the effect changes in response. Essentially, proofs are an immutable tool to aid research in its search for causal relationships.

We begin with a history of causal understanding then describe current tools for causal analysis, and move on to recent developments in causal analysis and how the dialogue between mathematical, medical and social sciences are pushing the boundaries of knowledge.

A brief, interdisciplinary history of causation

The tension between existing coherent causal theories (understanding of biological mechanism, domain expertise, etc.) and the results of models, built from observational data has a rich history. Understanding the nature of the current dialogue between theory and practice requires knowledge of the history of causal understanding as well as the state of existing methodologies. This knowledge base prompts discussion about pivotal new modelling techniques that allow for more nuanced representations of causal mechanisms.

The complexities of statistical models may sometimes obscure what scientists actually mean by cause. Helpfully, Cox and Wermuth identify three broad types of causality (**Cox and Wermuth, 1996**):

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

73

A. Causality as a statistical dependence which cannot be removed by alternative acceptable explanatory variables

B. Causality as inferred consequence of some intervention in the system

C. Causality as inferred consequence of some intervention in the system augmented by some understanding of a process or mechanism accounting for what is observed.

Questions of life and death prompted the earliest causal investigations. Several disciplines point to John Snow as one of the first scientists to frame causal questions systematically. John Snow was a medical doctor with an aptitude for mathematics. In the 1830s, Snow worked to understand what caused a devastating outbreak of cholera (**Snow, 1855**). At the time, the causal mechanism for cholera was unknown and Snow was sceptical about the prevailing miasma theory. After plotting on a map the number of outbreaks on each street and examining the counts, Snow noticed that the deaths clustered around the Broad Street water pump. However, no one working at the nearby brewery was getting ill. Further enquiry revealed that brewery workers drank beer during the lunch hour and not water from the Broad Street pump. His findings helped to prove that water contamination was a cause of cholera, drawing a causal conclusion from observational data. This is type B causality in the Cox and Wermuth framework.

Around the same time, Semmelweis investigated deaths in a hospital where he observed a decline in childbed fevers when doctors washed their hands with chlorinated lime after working on cadavers and before attending the maternity unit. His careful charting looks convincing today, but without the relevant germ theory developed and proved a few decades later through experimentation by scientists like Pasteur and Koch, Semmelweis's findings failed to find an audience. This unfortunate failure to communicate convincingly an unexpected causal link, type B causality in the Cox and Wermuth framework, underlines the fact that researchers expect to find explanations that cohere with their existing causal theories. The opposition Semmelweis met with after uncovering a new biological mechanism highlights just how difficult causal questions are to answer, particularly when the only information at our disposal is observational.

A few years later, economist George Yule began asking causal questions as he investigated the relationship between the changes in poverty rate with the proportion of public benefits (**Yule, 1895**). Controlling for confounders and multiple variables (Type A causality) proved pivotal to a multiple linear regression model, laying the foundation for much of the regression tools that proliferated among economists.

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

74

Phillip Wright sought to investigate the relationship between tax policy and demand and supply elasticity (**Wright, 1928**). His work on this subject introduced the concept of instrumental variables (IVs). In cases where a correlation exists between the explanatory variable of interest and the error term in the model, a variable that is correlated with the explanatory variable of interest, but independent the error term or outcome, may be added to the model. Then, by holding one variable constant and varying another, how the other changes can be used to infer relationships. For example, tobacco tax affects tobacco use but not health making it a candidate IV to investigate the causal link between smoking and health. Instrumental variables, another example of type A causality, have been a formative tool for addressing questions of causation in economics. Later developments in statistical theory would further augment the importance of IVs.

As causal questions gathered momentum in the medical, social, and mathematical sciences, psychologists questioned how humans understand cause. Kenneth Craik, pioneer of mental models, maintained that, 'If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way react in a much fuller, safer, and more competent manner to the emergencies which face it' (**Craik, 1944**). People construct internal models to represent the causal texture of the environment (**Tolman and Brunswik, 1935**). Understanding the nature of how humans comprehend cause is crucial to avoiding fallacies in the development of causal models and introducing error and uncertainty.

A critical advancement in causal thinking occurred in 1965, when epidemiologist Sir Austin Bradford Hill proposed a set of criteria for determining causal relationships from observational studies. Working together with physician Richard Doll, they had uncovered the causal link between smoking and lung cancer from observational data. Since a randomised controlled trial would be both unethical and infeasible, Bradford Hill and Doll instead had interviewed lung cancer patients about their smoking habits and exposure to other postulated causes (exposure to car fumes, tarmac dust and coal fire dust). In one of the earliest case-control studies, they matched the lung cancer patients to patients with carcinomas of the stomach or colon by age, sex, social class, and place of residence. The risk of developing lung cancer proved to be 50 times greater among patients who smoked 25 or more cigarettes a day when compared with non-smokers. This established a strong correlation, but in order to definitively establish the causal link, they undertook a prospective study of over 24,000 smokers and non-smokers among male medical

professionals aged over 35 (who smoked at the same rates as other occupations at the time). The causal link (type A and C causality) was demonstrated by a clear dose-response among smokers and a clear difference in rates of lung cancer between smokers and non-smokers. Causality was accepted when they published these preliminary findings three years later (**Doll and Bradford Hill, 1950**).

From these observational studies, Bradford Hill went on to establish a set of criteria for determining cause from observational data. The criteria require that a causal effect should demonstrate:

• Strength, a large effect size

• Consistency (reproducibility)

• Specificity to a particular population, site, disease

• Temporality, cause happens before effects,

• Appropriate biological gradient, dose response

• Plausible mechanism between cause and effect

• Coherence, agreement between observations and laboratory results

• Experimental evidence, where practicable

• Analogy, similarity to the effect of similar factors.

Since their development in 1965, the criteria have provided a hallmarks of causal links for medicine, epidemiology, and public health.

In the late 1980s and early 1990s the next important advance was the development of new types of graphical models, probabilistic models in which a graph expresses a conditional dependence structure between variables, and probability captures uncertainty by using appropriate distributions of values rather than point estimates. These uncover type A and C causality. Clive Granger proposed econometric time series models that defined a Granger cause (**Granger, 1988**) when the cause occurs prior to the effect and the cause has unique information about the future values of its effect. Rubin developed a potential outcomes framework, posing counterfactual questions about what would have been observed had different conditions prevailed. Judea Pearl and other statisticians and computer scientists explored the rich space of probabilistic graphical models that have been successfully applied to a vast array of applications, specifically Bayesian networks (BNs, see Figs. 1 and 2). Graphical models have now become ubiquitous and are typically one of the first tools used to address causal questions.

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

76

As network data becomes more readily available in the medical and social sciences, enabling the use of probabilistic graphical models, the debate about the importance of theory versus methodology is more pressing than ever. We will review these methods below before posing interdisciplinary questions at the boundary of current causal theory and methodology in the subsequent section.

Current tools of causal analysis

Current practices for determining causation vary across disciplines, in line with the history of each discipline. Randomised control trials remain the gold standard, especially in medicine, but where this is infeasible, scientists turn to discipline-specific tools to analyse observational data. Medical scientists tend to place particular importance on existing theories about causal pathways, in preference to allowing causal discovery algorithms to guide their experiment design. Since it is usually infeasible to measure every possible variable, current understanding about plausible causal explanations tend to drive experimental design.

## Bradford Hill criteria for study design

Epidemiology uses the Bradford Hill criteria to synthesise results from observational studies. For example, the Bradford Hill criteria have been used to make the case that sleep deprivation is a cause of obesity and of several chronic diseases, each criterion satisfied by different study designs (**Cappuccio *et al.*, 2010**). In the USA, as the average number of hours adults reported sleeping declined from 9.0 in 1910 to 6.8 in 2005, average BMI in the same population rose from 23.0 in 1910 to 26.9 in 2005, suggesting an association. A cross-sectional meta-analysis (**Cappuccio *et al.*, 2008**) showed an association between short duration of sleep and obesity prevalence (the proportion of cases in the population at a time point) in both children and adults, demonstrating the strength, specificity and consistency of association. A prospective study to see if exposure precedes outcome then determines the directionality. Recent work measured obesity incidence and showed that both children and adults with short sleep had an increased risk of developing obesity over time with the same order of magnitude. Carefully designed, short term randomised controlled trials of short and disturbed sleep determined a dose-response in key hormonal changes which replicated across people and reversed when sleep returned to normal. Observed changes in the levels of two hormones, leptin and ghrelin, which regulate appetite provide a plausible biological mechanisms for sleep deprivation causing obesity. Analogous results found plausible mechanisms for sleep deprivation causing diabetes, hypertension and coronary heart disease. (**Spiegel *et al.*, 2009, Broussard *et al.*, 2012; Cappuccio *et al.*, 2011; Leng *et al.*, 2015**).

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

77

In all of these studies, the effects are strong (large relative risks), consistent, show a temporal sequence, a dose – response, have biological plausibility and reversibility in controlled trial conditions (at least short-term), so under the Bradford Hill criteria we accept that poor sleep causes obesity and causes these chronic diseases. These criteria continue to guide study design in epidemiological research today. Understanding how these criteria are implemented is crucial for appropriate extra-disciplinary applications of statistical methods. This represents an area of growth for interdisciplinary work between statisticians and epidemiologists.

Causation as invariant statistical dependence in graphical models

Whilst the medical sciences are focused on uncovering the biological mechanism responsible for cause in a system in order to provide appropriate interventions, the mathematical sciences offer a powerful alternative to the traditional way of addressing this problem. Probabilistic graphical models offer a tremendous opportunity to inform experimental design alongside the qualitative considerations outlined above.

To address the questions of statistical dependence and consequences of interventions, scientists across disciplinary boundaries often use graphical models. Partially inspired by Wright's work on path diagrams (**Wright, 1934**), researchers began to use graphical models to depict the relationships between elements of a system. One of the most pervasive types of graphical model is the Bayesian network. These structures model problems as sets of nodes and directed edges without cycles. For instance, our system might be represented by the Bayesian network in Figure 1.
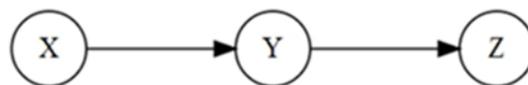
Figure 1: A sample Bayesian network. In this graph, X, Y and Z are random variables and the directed edges (arrows) represent the dependencies between the variables. This is a directed acyclic graph (DAG), since it is not possible to return to any node by following the directed edges. Formally, a BN is a directed acyclic graph and a set of independence statements.

Whilst it is tempting to interpret the arrows here as strictly causal, the mathematical interpretation of this graph is rather less strong. Under the critical Markov assumptions, it tells us that X and Z are independent given Y (**Markov, 1954**). Missing edges in graphs (such as that between X and Z) demonstrate independences in the graph—our starting point for determining what doesn't cause what in a graph. In fact, each graph belongs to a class of equivalent graphs that encode the same conditional independence relationships.

In our example, the Bayesian networks in Figure 2 are all equivalent because they all represent X and Z as conditionally independent. Bayesian networks also fulfil the Faithfulness Assumption, which means that all of the necessary conditional independence statements are encoded by the network. That is, there are no additional, context-specific independences necessary to model the system (**Meek, 1995**; **Spirtes *et al.*, 2000**).
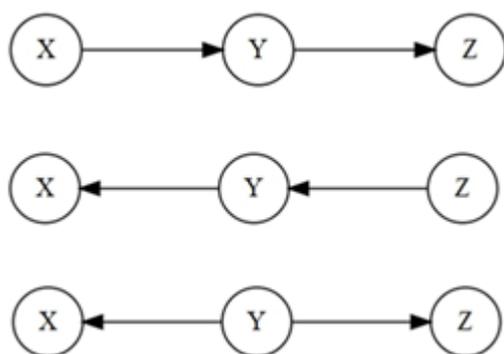

Figure 2: Equivalent Bayesian networks, all encoding that X is independent of Z given Y.

To determine what Pearl terms 'genuine cause' our model must admit an instrumental variable (Figure 3) to identify which causes are invariant across the class of equivalent graphs. A Bayesian network is truly causal when each of the nodes is invariant to marginalisation. That is, forcing a variable node to take a particular value has the same effect on the other nodes as if the variable had taken that value naturally.
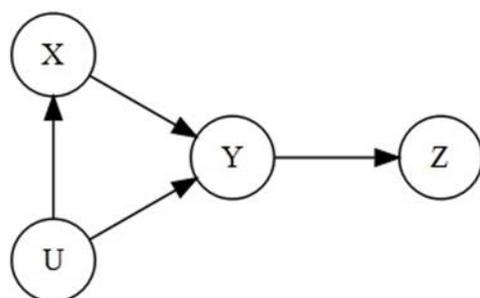

Figure 3: Bayesian network encoding that Z is independent of both X and U, given Y. Z is an instrumental variable. i) Z is associated with the treatment X. ii) Z is independent of the unobserved confounding factors affecting our treatment and outcome (X and Y, respectively). iii) Z is independent of Y given X and the unobserved confounders.

Using these fundamental assumptions, Bayesian networks that are invariant to marginalisation can be used to determine mathematically causal relationships within a system. These networks can then be used to estimate the effect of proposed interventions using Pearl's Do-Calculus (**Pearl, 2009**), and so can form the basis of policy decision support.

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

79

The promise of Bayesian networks has catalysed research into algorithms to find network structure from data and causal relationships within a network (**Entner, 2013**). Causal discovery algorithms have expedited the disciplinary reach of these methods, and as more and more data becomes available, the social sciences are beginning to leverage these methods. However, their usefulness is limited unless used alongside domain expertise and qualitative considerations, such as that used in the Bradford Hill criteria.

Recent advancement in causal analysis

The collaboration between quantitative and qualitative approaches to causation represents a key area for research growth, providing methods to combine data, expert knowledge and plausible assumptions to reach causal conclusions. Mathematicians are working to refine the assumptions of graphical models, remedy limitations of scope in existing methodology and define new classes of models that may be better suited to specific causal questions raised in an array of disciplines.

Refining model assumptions

When using observational data the question arises how we can compute the causal effect of one variable on another from data obtained from passive observation without interventions? By using a graph to represent the problem, this becomes a graph-theoretic problem. Pearl (**2016**) introduced a back-door criterion, identifying which variables should be conditioned on when investigating a causal relationship between other variables.

Bayesian networks can legitimately be used as long as the model meets the Markov and faithfulness assumptions. The faithfulness assumption strengthens the inferences we can draw in some practical applications (e.g. **Chen *et al.*, 2007**). A live area of research seeks to relax these and other assumptions in robust ways to obtain more flexible notions of causal inference.

One recent example of a method for relaxing the faithfulness assumption for when treatment and outcome are confounded is the witness protection program protocol developed by Silva and Evans (**2016**). This provides ways to find a set of variables that allow a witness variable to be used as an instrumental variable to give bounds on the average causal effect. This also allows us to differentiate between strong directed effects and strong active paths and thus more nuanced definitions of causation. In this way, the new witness variable bridges back-door adjustment and the IV adjustment via the faithfulness assumption.

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

80

The importance of assumptions in mathematical and modelling cannot be overstated: the real purpose of causal discovery methods is not to provide neat answers, but rather to demonstrate that observational data is compatible with more tentative answers. For this it may be necessary to devise new models

Defining new models: the chain graph

The complex, rich and diverse tasks of the statistician include understanding research questions in other fields, designing empirical studies, evaluating models and methods of analysis, and interpreting evidence in data and results of statistical analyses. Understanding a problem statistically requires thorough investigation of the context, response variables of interest, regressors, and intermediate or mediating variables.

However, naïve or inappropriate use of statistical methodology can lead to indefensible conclusions and studies which fail to replicate the same results with appropriate alternate data sets. Failure to replicate a studied effect calls into question the hypothesised causal relationships. From a statistical perspective, some methods may not permit replication. Consequently, some measures of dependence are inappropriate if replication under stated conditions is a purpose of the research, such as in medical applications. For example, applying multivariate methods to several binary variables will not give replication when the context conditions of studies differ strongly. To address this, Wermuth and Marchetti (**2017**) demonstrate that replicable results are permitted by well-fitting, mean zero Ising models (**Ising, 1925**).

Another difficulty is that, in a given study, joint responses may be needed to properly capture effects of interventions. For instance, a medication to treat high blood pressure affects both systolic and diastolic blood pressure simultaneously, so it is not appropriate to model them as occurring sequentially. Hence, models should include joint responses whenever no order is plausible for several responses which remain related after their regressions on important explanatory variables.

To circumvent the limitations of the DAG models (Figure 1) to capture joint responses, Wermuth and Cox (**2013**) suggest regression graphs, a chain graph structure that encodes the ordering of joint responses by blocking together variables of the same type (response, intermediate, explanatory, background) and that represent conditional independences between them. These graphs are particularly adept at representing joint responses, (**Cox and Wermuth, 1993; Drton, 2009; Fallat *et al*., 2017**).

Interdisciplinary efforts towards careful definition of the problem context, possible regressors, and joint responses motivates advancements in probabilistic graphical models.

Defining new classes of models as series of events

While the Bayesian network is a powerful tool for causal models, it is not always appropriate. Often, causal claims may be presented as a narrative of sequential events which can be described mathematically by an event tree. These can be transformed to a new class of graphical model, the chain event graph (CEG) which admits a unique causal algebra. Using algebraic statistics, we can extend the machinery of Bayesian networks to other classes of graphical mode in the following way.

In an undirected graph, the directionality of the relationships between variables represented by an undirected edge is ambiguous, so no causal structure can be inferred. In a directed graph, such as a Bayesian network (BN), we have potential causal relationships given by the set of possible collections of conditional independence statements that describe the data. Causal discovery algorithms score possible BNs according to how well they fit the data. Often, these discovered graphs are then reverse-engineered to infer causal implications and estimate causal effects, assuming graph is the truth. Although the output of the causal discovery algorithms is often taken to be causal, Spirtes and Pearl argue that such output alone is not sufficient to deduce a cause (**Spirtes, 2000**; **Pearl, 2009**) since there is an entire class of statistically equivalent graphs which can be represented by the same essential graph, having only the directed edges common to all graphs in the class. These are all candidates for truly causal relationships. Pearl further defines a genuine cause in a graph as a random variable that has an associated instrumental parent within its essential graph.

This idea can be expanded beyond BNs to other classes of graphical models. In particular, we now have a suite of tools that allow us to fully explore chain event graphs (CEGs). Cowell and Smith (**2014**) develop causal discovery techniques to find the best fitting CEG from data. Thwaites *et al.* (**2010**) demonstrated how the causal hypotheses of a CEG offer a profound flexibility. These classes have been extended even further using computer algebras (**Görgen, 2017**). Görgen and Smith (**2017**) define the statistical equivalence classes of staged trees, identifying the set of possible representations. Potential causal directionality on variables deduced from quaternion relationships can be found from algebraic features shared by all elements in an equivalence class. This is analogous to essential graphs. As shown by Collazo *et al.* (**2018**), this new class of graphs accommodates a much richer space of causal hypotheses.

Impact on the social sciences

Joint responses, asymmetric and sequential relationships seen in the medical and social sciences have driven the development of new mathematical models to explore them. Mathematical models are also used to explore mental models.

For psychologists, one open question regarding causation is how do people update their mental models of cognition? Causal judgements help us learn to predict and control our world, build causal models, to reason about evidence, and to determine how we attribute responsibilities to ourselves or others, especially with respect to legal or medical evidence. Causal judgements also affect our temporal beliefs; it has been shown that these mental models can even override our perception of the order in which things happen (**Bechlivanidis and Lagnado, 2016**). When learning about a system, we may choose interventions to target local uncertainty in a model (**Bramley *et al*., 2015; 2017**). The metaphor of Neurath's ship describes how experience is believed to update mental causal models. 'We are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support. In this way … the ship can be shaped entirely anew, but only by gradual reconstruction' (**Quine, 1960**). In the absence of knowing all possible models, a current mental model shifts slowly in the light of evidence, adapting with local changes.

Current studies are underway to investigate how people learn graphical models. While the development and proliferation of Bayesian network methods proved a useful tool for psychologists to articulate new models of cognition, their work in turn has profound insight for how statisticians communicate their results to convey understanding to non-specialists. This is particularly important when probabilistic graphical models are used to underpin decision support mechanisms designed to evaluate alternate courses of action (**Smith, 2010**; **Smith *et al*., 2015**; **Barons *et al*., 2018**).

Economists and medical researchers share a common interest in determining the role and importance of theory in these new causal experiments. Some economists (**Deaton and Cartwright, 2016**; **Heckerman *et al*., 1995**) advocate that there can be no causation without theory. This ties in to the Bradford Hill criterion—that we expect any potentially valid causal explanation not to contradict known theories. Medical professionals are concerned about the increasing reach of causal discovery algorithms to suggest new biological mechanism. For example, in Hill *et al*. (**2012**), dynamic Bayesian networks were used to infer the protein signalling network structure in a breast cancer cell line. This

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

83

generated testable hypotheses, which were then independently validated using targeted inhibition, improving knowledge about breast cancer.

The debate about the importance of theory is a rich opportunity for computational scientists to develop models more attune to relevant theories. Discovering reproducible, robust causal results demands working alongside other disciplines to further the symbiotic relationship between the mathematical and social sciences.

Discussion

Causal theory and techniques to determine causation from observational data have an uneasy alliance. The mathematical sciences have produced powerful statistical models and machine learning techniques that have rendered causal discovery techniques accessible to a new range of social sciences. Economists and psychologists are using these new methods to ask more in-depth questions about drivers in economics and cognition. Causal discovery techniques discover plausible biological mechanisms with increasing rapidity.

Mathematical methods are inevitably subject to the underlying mathematical assumptions on which their foundational proofs rely. Ideally, causal analysis across disciplines works in tandem, with model results guiding further discovery in the medical or social sciences. In turn, practitioners in medical and social sciences may expose the limitation of current methodology, which in turn prompts mathematical developments to create bespoke models. The tension between existing theories about causal mechanisms can be a productive conversation rife with new research opportunities.

Acknowledgments

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

84

References

Barons, M. J. and P. I. Chleboun, (2015), 'Exchanges' - Conversations with… Professor Martin Hairer. *Exchanges: The Warwick Research Journal*, Vol 3 No 1. Retrieved from: https://exchanges.warwick.ac.uk/article/view/123

Barons, M. J., Wright S. K. and J. Q. Smith, (2018), 'Eliciting Probabilistic Judgments for Integrating Decision Support Systems', in: Dias LC, Morton A, Quigley J (eds), *Elicitation: The Science and Art of Structuring Judgement.* New York, Springer, Chapter 17 pp 445-478

Bechlivanidis, C., and D. A. Lagnado, (2016), Time reordered: Causal perception guides the interpretation of temporal order. *Cognition*, 146, 58–66. http://doi.org/10.1016/j.cognition.2015.09.001

Bramley, N. R., Dayan, P., Griffiths, T. L., and D. A. Lagnado, (2017), Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, 124(3), 301–338. Retrieved from http://psycnet.apa.org/journals/rev/124/3/301/

Bramley, N. R., Lagnado, D. A., & M. Speekenbrink, (2015), Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 708–731. http://doi.org/10.1037/xlm0000061

Broussard, J. L., Ehrmann, D. A., Van Cauter, E., Tasali, E., and M. J. Brady, (2012), Impaired Insulin Signaling in Human Adipocytes After Experimental Sleep Restriction A Randomized, Crossover Study. *Annals of Internal Medicine*, 157(8), 549–557. http://doi.org/10.7326/0003-4819-157-8-201210160-00005

Cappuccio, F., Miller, M. A. (Physician), and S. W. Lockley, (2010), *Sleep, health, and society : from aetiology to public health*. Oxford, UK. Oxford University Press.

Cappuccio, F. P., Cooper, D., Delia, L., Strazzullo, P., and M. A. Miller, (2011), Sleep duration predicts cardiovascular outcomes: A systematic review and meta-analysis of prospective studies. *European Heart Journal*, 32(12), 1484–1492. http://doi.org/10.1093/eurheartj/ehr007

Cappuccio, F. P., D'Elia, L., Strazzullo, P., and M. A. Miller, (2010), Quantity and quality of sleep and incidence of type 2 diabetes: A systematic review and meta-analysis. *Diabetes Care*, 33(2), 414–420. http://doi.org/10.2337/dc09-1124

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

85

Cappuccio, F. P., Taggart, F. M., Kandala, N.-B., Currie, A., Peile, E., Stranges, S., and M. A. Miller (Physician), (2008), Meta Analysis of Short Sleep Duration and Obseity in Children and Adults. *Sleep*, 31(5), 619-26.

Chen, L. S., Emmert-Streib, F., and J. D. Storey, (2007), Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biology*, 8(10), R219. http://doi.org/10.1186/gb-2007-8-10-r219

Cox, D. R., and N. Wermuth, (1993), Linear Dependencies represented by Chain Graphs. *Statistical Science*, 8(3), 204–83.

Cox, D. R., and N. Wermuth, (1996), *Multivariate Dependencies. Models, Analysis and Interpretation*. London, UK. Chapman & Hall.

Cowell, R.G. and J. Q. Smith, (2014), Causal discovery through MAP selection of stratified chain event graphs. *Electronic J of Statistics*, 8, 965 – 97

Collazo, R.A., Gorgen, C and J. Q. Smith, (2018), *Chain Event Graphs*, London, UK. Chapman Hall

Craik, K. J. W., (1944), *The Nature of Explanation*. London: Cambridge University Press. p. 61

Deaton, A., and N. Cartwright, (2016), *Understanding and Misunderstanding Randomized Controlled Trials*. Cambridge, MA. National |Bureau of Economic Research Working Paper 22595 http://doi.org/10.3386/w22595

Doll, R. and A Bradford Hill, (1950). Smoking and Carcinoma of the Lung *Br Med J*, 2 (4682), 739-748 doi: https://doi.org/10.1136/bmj.2.4682.739

Drton, M. (2009), Discrete chain graph models. *Bernoulli*, 15(3), 736–53. http://doi.org/10.3150/08-BEJ172

Entner, D., Hoyer, P. O., and P Spirtes, (2013), Data-driven covariate selection for nonparametric estimation of causal effects. *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, 31, 256–64.

French, S. J., (2015), Cynefin: uncertainty, small worlds and scenarios' *Journal of the Operational Research Society*, 66(10): 1635-45. https://doi.org/10.1057/jors.2015.21

Fallat, S., Lauritzen, S., Sadeghi, K., Uhler, C., Wermuth, N., and P. Zwiernik, (2017), Total positivity in Markov structures. *Annals of Statistics*, 45(3), 1152-1184. http://doi.org/10.1214/16-AOS1478

Görgen, C., and J. Q. Smith, (2015), Equivalence Classes of Staged Trees. *arXiv*, 1512.00209.

Granger, C. W. J. (1988), Causality, cointegration, and control. *Journal of Economic Dynamics and Control*, 12(2–3), 551–9. http://doi.org/10.1016/0165-1889(88)90055-3

Heckerman, D., Geiger, D., and D. M. Chickering, (1995), Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*, 20(3), 197–243. http://doi.org/10.1023/A:1022623210503

Hill SM, Lu Y, Molina J, Heiser LM, Spellman PT, Speed TP, Gray JW, Mills GB, and S. Mukherjee, (2012) Bayesian inference of signalling network topology in a cancer cell line. *Bioinformatics*, 28(21): 2804-10. https://doi.org/10.1093/bioinformatics/bts514

Ising, E., (1925), Beitrag zur Theorie des Ferromagnetismus, *Z. Phys*., 31: pp253–8, doi:10.1007/BF02980577.

Kuttz, C.F. and D. J. Snowden, (2003), The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Systems Journal*, 42(3), 462-83.

Leng, Y., Cappuccio, F. P., Wainwright, N. W. J., Surtees, P. G., Luben, R., Brayne, C., and K. T. Khaw, (2015), Sleep duration and risk of fatal and nonfatal stroke A prospective study and meta-analysis. *Neurology*, 84(11), 1072–9. http://doi.org/10.1212/Wnl.0000000000001371

Markov, A. A., (1954), Theory of Algorithms. *Trudy Mat*. *Inst. Steklov*., 42, Acad. Sci. USSR, Moscow–Leningrad, pp3–375

Meek, C., (1995), Strong Completeness and faithfulness in Bayesian Networks in *Uncertainty in Artificial Intelligence* 11 (P. Besnard and S. Hanks eds) Morgan Kaufmann pp 411-8

Pearl, J. (2009). *Causality* (2nd ed.). New York, US: Cambridge University Press.

Pearl, J., Glymour, M., and N. P. Jewell, (2016), *Causal inference in Statistics*, West Sussex, UK, Wiley.

Quine, W, (1960), *Word and Object*. Cambridge, Mass: MIT Press. P3.

Semmelweiss I., (1867), *The Etiology, Concept, and Prophylasis of Childbed Fever*. Translated by K. C. Carter, University of Wisconsin Press, 1983.

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

87

Silva, R., and R. Evans, (2016), Causal inference through a witness protection program. *The Journal of Machine Learning Research*, 17(1), 1949-2001.

Sloman, S. A., and D. Lagnado, (2015), Causality in Thought. *Annual Review of Psychology*, 66(1), 223–47. http://doi.org/10.1146/annurev-psych-010814-015135

Smith, J.Q., (2010) ' Bayesian Decision Analysis: Principles and Practice' Cambridge, UK. Cambridge University Press

Smith, J.Q., Barons M. J. and M. Leonelli, (2015), Decision focused inference on networked probabilistic systems: with applications to food security. *JSM Proceedings*, Seattle, Washington: American Statistical Association. pp 3220-33.

Snow J., (1855), *On the Mode of Communication of Cholera*. London Churchill. Reprinted by Hafner, New York, 1965.

Spiegel, K., Tasali, E., Leproult, R., and E. Van Cauter, (2009), Effects of poor and short sleep on glucose metabolism and obesity risk. *Nature Reviews Endocrinology*. 5, 253-61. http://doi.org/10.1038/nrendo.2009.23

Spirtes, P., Glymour, C. and R. Scheines, (2000), *Causation, Prediction and Search*. Cambridge, UK. Cambridge University Press.

Steptoe, A., (2006), Sleep Duration and Health in Young Adults. *Archives of Internal Medicine*, 166(16), 1689-92. http://doi.org/10.1001/archinte.166.16.1689

Thwaites, P. Smith, J.Q. and E. Riccomagno, (2010) Causal Analysis with Chain Event Graphs. *Artificial Intelligence*, 174, 889-909

Tolman, E. C., and E. Brunswik, (1935), The organism and the causal texture of the environment. *Psychological Review*, 42(1), 43–77. http://doi.org/10.1037/h0062156

Wermuth, N., and D. R. Cox, (2013), Concepts and a case study for a flexible class of graphical markov models'. *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*, 331–50. http://doi.org/10.1007/978-3-642-35494-6_20

Wermuth, N., and G. M. Marchetti, (2017), Generating large Ising models with Markov structure via simple linear relations'. Retrieved from http://arxiv.org/abs/1704.01649

Woodward, J., (2003), *Making Things Happen: A Theory of Causal Explanation*. (Vol. 14). Oxford, UK. Oxford University Press. P220 http://doi.org/10.1080/10705510709336742

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

88

Wright, P. G., (1928), *The Tariff on Animal and Vegetable Oils*. Institute of Economics Investigations in International and commercial policies. New York. The Macmillan Co. Retrieved from http://dspace.gipe.ac.in/xmlui/bitstream/handle/10973/24323/GIPE-079282-Contents.pdf?sequence=2&isAllowed=y

Wright, S., (1934), The method of path coefficients. *Annals of Mathematical Statistics*. 5 (3), 161–15. http://doi:10.1214/aoms/1177732676

Yule, G. U., (1895), On the Correlation of Total Pauperism with Proportion of Out-Relief. *Economic Journal*, 5(20), 603–11.

Yule G. U., (1899), An investigation into the causes of changes in pauperism in England, chiefly during the last two intercensal decades. *J. Roy. Statist. Soc.*, 62, 249-95.

Barons & Wilkerson. Exchanges 2018 5(2), pp. 72-89

89