



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Using interventions to discover quantum causal structure

Sally Shrapnel

BMedSci, MBBS, MSc (Eng) DIC (Dist)

*A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2016
School of Historical and Philosophical Inquiry*

Abstract

Identifying causal relations plays an indispensable role in science, economics, medicine and everyday life. Interventionist causation associates the discovery of such relations with the possibility of manipulation or intervention. A causes B , if manipulating A in the right way can bring about changes in B . The main technical tool, the causal model, can be seen as a device for telling us how and why certain actions are more effective than others. Such models have found application in almost every scientific discipline.

However, despite several decades of sophisticated analysis and a wealth of technical results, interventionist causal methods are known to fail in the quantum case. The quantum domain, with its peculiar, counterintuitive features, is widely considered an inhospitable environment in which to apply our usual causal discovery tools.

This causal skepticism presents us with a puzzle. Physicists are now able to manipulate and control quantum systems, and the design and construction of quantum technology relies on an ability to identify when one strategy will work and another will fail. From the perspective of interventionist causation, it seems that a formal account of quantum causal modelling ought to be possible.

The main contribution of this thesis is to provide an extension of interventionist methodology to the quantum case. In Chapter 1, I introduce interventionist causation from the perspective of two giants of the field: Judea Pearl and James Woodward. The Causal Markov Condition (CMC) and Faithfulness assumptions are presented, and I draw particular attention to the discovery of causal structure. I touch on some of the well-known problems with interventionist causation *per se*.

In Chapter 2, I put this methodology to work in a quantum mechanical context. Following tradition, we examine one of the Bell experiments in order to determine how the classical interventionist framework fails. Noting that the usual response is to demand that one drops either the CMC or Faithfulness, I suggest that neither approach is particularly satisfactory and argue for an alternative. I discuss some reasons why the interventionist is forced to pursue this different path.

In Chapter 3, I motivate the possibility of an interventionist account of quantum causation. I first look at how physicists represent quantum technology in order to make interventionist and counterfactual inferences: via the use of quantum circuit diagrams. Using three examples I demonstrate the manner in which these diagrams can be used to facilitate causal inference.

In Chapter 4, I examine some recent attempts at characterising quantum causal models due to work from the quantum foundations community. I present three such accounts and explain why I endorse a fourth, the focus of the next chapter.

In Chapter 5, I present a new framework that allows for the discovery of interventionist quantum causal structure. I finish the thesis with a summary and some suggestions for further work.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

- (i) Shrapnel, S. (2014). Quantum Causal Explanation: Or, Why Birds Fly South. *European Journal for Philosophy of Science*, 4: 409–423.
- (ii) Shrapnel, S. (2015). Discovering quantum causal models. Submitted to *British Journal for Philosophy of Science*, 2015, **(currently under review)**.
- (iii) Costa, F and Shrapnel, S. (2016). Quantum Causal Modelling. *New Journal of Physics*, 18 (6), 063032.

Publications included in this thesis

- (i) Shrapnel, S. (2014). Quantum Causal Explanation: Or, Why Birds Fly South. *European Journal for Philosophy of Science*, 4: 409–423. Incorporated as Appendix.

Contributions by others to the thesis

Professor Phil Dowe has made important contributions to this thesis by providing advice and feedback on several drafts of my published paper (detailed above), and also on the drafts of this thesis. Professor Gerard Milburn has made important contributions by providing general advice on open quantum systems and detailed advice on the examples of Chapter 3: the Laser Gravitational Wave Observatory, the nitrogen-vacancy diamond and the avian magneto-receptive mechanism. Professor Milburn also provided helpful comments on drafts of my published paper (detailed above) and drafts of this thesis. Dr. Peter Evans contributed by providing helpful comments on final drafts of this thesis.

The process matrix approach to quantum causal modeling, discussed in Chapter 5 grew out of many discussions with Fabio Costa, one of the original authors of the process matrix formalism. The results of these discussions are contained in our joint paper (detailed above), and the contribution statement to this paper can be found below.

Contributor	Statement of contribution
Sally Shrapnel	Conceptual work (80%) Technical work (20%) Wrote the paper (40%)
Fabio Costa	Conceptual work (20%) Technical work (80%) Wrote the paper (60%)

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

I would like to thank my advisory team. Without the guidance and support I have received from Gerard Milburn, Phil Dowe and Peter Evans, this thesis would not have come into existence. I am very grateful to Pete for stepping in at the eleventh hour and helping me with the final stages, for his enthusiasm and encouragement throughout and for his open-door policy when it came to needing help. Thanks also to Phil for sharing his knowledge regarding the metaphysics of causation, his exceptional red-herring detector and superb editorial skills. Appreciation and thanks also to Gerard for his clear and patient explanations of all matters quantum, his generosity, professionalism, intellect and consistently good humour. I could not have asked for a better role model.

For funding, I gratefully acknowledge assistance from a UQ PhD research scholarship.

I would like to thank the following UQ philosophy and physics students and postdocs for useful discussions throughout various stages of the PhD: Milena Ivanova, Matt Farr, Alex Varlakova, Dave Kinkead, Christina Giarmatzi, Magdalena Zych and Martin Ringbauer. Particular thanks to Fabio Costa: who shared my enthusiasm for causal modelling and also taught me numerous useful quantum foundational facts. Thanks also to Drew Ringsmuth for encouraging me to use LaTeX, for extolling the virtues of Mike and Ike, and for keeping me laughing throughout.

Thanks to the participants of the 2013 Australasian Association of Philosophy Conference, the 2014 Durham Summer School on Emergence and the 2015 Causal Power of Information in a Quantum World Conference, for many illuminating and enjoyable discussions.

Thank you to the following people for personal support and encouragement during various stages of the PhD: Eva King, Nina Sayer, Kylie and Scott Deegan, and Pam and Rob Shrapnel.

Finally, special thanks to Innes, John and Alexander for their support and love.

Keywords

Quantum causal models, interventionist causation, quantum causation, causal modelling, probabilistic graphical models, quantum causal structure, quantum machine learning

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 220206, History and Philosophy of Science (incl. Non-historical Philosophy of Science) 60%

ANZSRC code: 020603, Quantum Information, Computation and Communication 20%

ANZSRC code: 020699, Quantum Physics not elsewhere classified 20 %

Fields of Research (FoR) Classification

FoR code: 2202, History and Philosophy of Specific Fields, 60%

FoR code: 0206, Quantum Physics 40%

Contents

List of figures	xiii
Introduction	1
1 Interventionist Causation	5
1.1 Introduction.	5
1.2 Pearl's Causal Bayesian Networks	8
1.2.1 Formalism	8
1.2.2 Interpretation	13
1.2.3 Causal Mechanisms	17
1.2.4 Type vs Token causation	19
1.2.5 Discovery Algorithms	19
1.2.6 Dependence separation	20
1.2.7 The Bigger Picture	22
1.2.8 Summary	25
1.3 Woodward's causal models	26
1.3.1 Methodology	26
1.3.2 Philosophical Worries	30
1.3.3 Circularity	30
1.3.4 Causal Markov Assumption	30
1.3.5 Faithfulness	31
1.3.6 General objections	32
1.3.7 A cautionary note	33
1.4 Summary	35

2	Interventionist Causation and Quantum Mechanics	37
2.1	Introduction	37
2.2	The Bell experiments.	38
2.2.1	The CHSH inequality	43
2.3	What gives? Causal Markov vs Faithfulness	47
2.3.1	Relinquishing the Causal Markov Condition	48
2.3.2	Relinquishing Faithfulness	51
2.3.3	Must we choose?	53
2.3.4	Pearl and Woodward on Bell’s violations	53
2.3.5	Using interventions to discover causal relations	55
2.4	Alternative explanations	57
2.5	Summary	58
3	Motivating Quantum Causal Models	59
3.1	Introduction	59
3.2	Engineered quantum systems	60
3.2.1	Quantum Circuits	62
3.3	Quantum Circuit Examples	65
3.3.1	Quantum optical interferometry	67
3.3.2	The Nitrogen-Vacancy Diamond	70
3.3.3	The European Robin	73
3.3.4	Conclusions	77
4	Quantum Causal Models: the state of play	79
4.1	Quantum Causal Networks	80
4.1.1	Intervention	83
4.1.2	Causal Discovery	84
4.1.3	Classicality	85
4.2	A Graph Separation theorem for Quantum Causal Models	85
4.2.1	Interventions	90
4.2.2	Classical Limit	90
4.3	Generalised Bayesian Networks	91

4.3.1	Generalised operational probabilistic theories	91
4.3.2	The Generalised Bayesian Network	95
4.3.3	D-separation	98
4.3.4	Causal Discovery	99
4.3.5	Classical limit	99
4.4	Summary	99
5	Markovian Quantum Causal Modelling	101
5.1	Introduction	101
5.2	Quantum variables and values	103
5.3	Process matrices and the generalised Born rule	106
5.4	Three node quantum causal models	115
5.5	Quantum Interventions	118
5.5.1	Quantum Direct Causation	119
5.6	Quantum Causal Model	120
5.6.1	Quantum Causal Markov Condition	122
5.7	Causal Discovery using Hilbert-Schmidt basis	123
5.7.1	Quantum fine-tuning	124
5.7.2	Effective and ineffective quantum strategies	125
5.8	The emergence of classical causal structure.	127
5.9	Bell’s objections	131
5.9.1	Universality	133
5.10	Conclusions	134
	Conclusions	137
	Bibliography	140
	Appendices	153
	App. A: Quantum Causal Explanation: or, why birds fly south	153

List of Figures

1.1	Three node DAG assuming causal sufficiency.	12
1.2	Unmutilated DAG	15
1.3	Mutilated DAG.	15
1.4	Three node DAG with latent variables	21
1.5	Three node DAG with collider	22
2.1	A possible DAG for the Bell experiment	44
2.2	Allowing for Superdeterminism	49
2.3	Allowing for Superluminal causes	50
3.1	General Quantum Circuit	64
3.2	Mach-Zender Interferometer	67
3.3	Quantum Circuit for LIGO	69
3.4	Quantum circuit for NV diamond.	73
3.5	Quantum circuit for avian magnetoreception.	76
4.1	Quantum causal network	81
4.2	Quantum Circuit DAG	87
4.3	Compositional elements	93
4.4	Test from trivial input to itself	93
4.5	A preparation-test, transformation and then observation-test.	94
4.6	GDAG for Bell experiment.	98
5.1	Spatio-temporal region.	105
5.2	Generalised circuit diagram	108
5.3	Bipartite entangled state: common cause.	116

5.5	Quantum common cause structure.	116
5.4	Direct cause structure	116
5.6	Quantum direct cause structure.	117
5.7	Markov quantum causal model.	121
5.8	Labelled node.	128
5.9	Mach-Zender Interferometer	159
5.10	Causal digram for radical pair model	168

Introduction

The aim of this thesis is to present an interventionist account of causation that is valid for both classical and quantum phenomena. Interventionist causation associates causal relations with the possibility of manipulation or intervention. The main technical tool, the causal model, can be seen as a guide to action, a device for telling us how and why certain strategies are better than others. For example, knowing that the mercury level and the onset of snow are correlated does not tell us whether turning the thermometer upside down will prevent snow. It is *causal* information that allows us to determine the right action in this situation.

Interventionist causation comes with a well-defined technical toolbox. It is generally considered that these tools fail to identify causal relations in the quantum case and many believe that interventionist techniques cannot be extended to the quantum domain (Glymour (2006), Spirtes et al. (2000), Pearl (2009), Koller and Friedman (2009)).

This view, however, presents us with a puzzle. It seems that we *are* able to make distinctions between effective and ineffective strategies in situations that involve quantum systems. Physicist can now exploit quantum phenomena to achieve new ends. They use models, graphical representations and explanations that describe new quantum technology and facilitate causal inference. They help scientists predict, manipulate and explain. They help identify when one strategy will be effective and when another will fail. For the interventionist, what can these models be, if not causal?

Starting with this very simple question, it is surprising how far one can go. Insofar as interventionist causation is about distinguishing between effective and ineffective strategies, it is possible to present a theory that captures this aspect of the quantum domain. One advantage of such an approach is that it is possible to remain metaphysically neutral and avoid the usual difficulties that come with quantum causation. Non-locality, the measurement problem and interpretational issues can be put on hold. Armed with a clear formalism, it is then possible

to assess the metaphysical implications quantum phenomena present for our understanding of causation.

The main contribution of this thesis is that it provides an extension of interventionist methodology to the quantum case. The formalism presented is mathematically consistent and recovers classical causal modelling structure in an appropriate limit. It represents an important challenge to the orthodoxy that one cannot discover interventionist quantum causal relations. Additionally, it provides an introduction to the very new field of quantum causal modelling, considers rival accounts and provides several motivating reasons to believe such an account ought to be possible.

There are, however, many philosophers who believe that physics in general is inhospitable to the existence of causal relations. Problems include the fact that the variables of fundamental physics are not coarse-grained (Woodward, 2007), that physical theories are time symmetric (Farr and Reutlinger (2013), Russell (1912)) and that fundamental physics aims at universality. The general picture for such “Russellians” is that physics gives us fundamental laws that tell us how the state of the universe evolves from one moment to the next. Thus, whilst plausibly relevant in certain *applied* situations (so the thinking goes), there is no need for causal notions to feature in our most fundamental physical theories.

Whilst I believe interesting counterarguments to this position have recently been presented (Frisch (2014a) and Ismael (forthcoming)), it is not my aim to mount a systematic attack on the position. Insofar as the Russellian perspective is an ontological claim (there is no place for causal relations in the ontology of fundamental physics), it is not my aim to deny such a view. If, however, one is referring to an eliminativist position that claims casual reasoning to be entirely absent, or somehow useless, in the actual practice of physics, then there is much in this thesis to dispute this. Ultimately, I leave the metaphysics to the reader: one can regard the quantum causal models I present in Chapter 5 simply as representational devices that give us nothing more than a handle on the practical advantages of causal reasoning involving quantum systems.¹ Alternatively, one can take the view that the pragmatic value of such characterisations should not be so easily dismissed.²

Additionally, I do not wish to suggest that interventionist causation is the *only* account of

¹This is sometimes called the “Republican” view. “Causal republicanism is the view that although the notion of causation is useful, perhaps indispensable, in our dealings with the world, it is a category provided by neither God nor physics, but rather constructed by us.” (Price and Corry (2007)) [p2]

²I provide two small, speculative points in Section 1.3.7 to support this alternative view.

causation with the potential to bear philosophical fruit. Rather, my chief aim is to assess to what extent we can extend interventionism to the quantum case.

A brief outline of the thesis is as follows: in Chapter 1, I introduce interventionist causation from the perspective of (Pearl, 2000) and (Woodward, 2003). The Causal Markov Condition (CMC) and Faithfulness assumptions are presented, and I draw particular attention to the discovery of casual structure. Finally, I touch on some of the well known problems with Interventionist causation.

In Chapter 2, I put this methodology to work in a quantum mechanical context. Following tradition, we examine one of the Bell scenarios in order to determine how the classical interventionist framework fails. We first briefly review standard approaches from outside the causal modelling literature, and adopt the view that Bell's theorem presents one with a choice between giving up on locality or on causal explanation. Using the interventionist causal formalism of Chapter 1, we characterise the manner in which this account fails to explain the Bell correlations. Noting the usual response is to advocate that one gives up either the Causal Markov condition or Faithfulness, we see that neither avenue is particularly attractive.

In Chapter 3, I motivate the *possibility* of an interventionist account that can accommodate quantum phenomena. I first look at how physicists represent quantum technology in order to make interventionist and counterfactual inferences: via the use of quantum circuit diagrams. Using three examples I demonstrate the manner in which these diagrams can be used to facilitate causal inference.

In Chapter 4, I examine some recent attempts at characterising quantum causal models. Due to the success of Pearl's causal modelling methods, the quantum foundations community has recently begun to examine the possibility of providing a quantum analogue. I introduce three such accounts and explain why I endorse a fourth, the focus of the next chapter.

In Chapter 5, I present a new framework that allows for the discovery of interventionist quantum causal structure, based on the process matrix formalism of Oreshkov et al. (2012). I finish up the thesis with some conclusions and suggestions for further work.

Chapter 1

Interventionist Causation

In this Chapter, we look at two different interventionist accounts (Pearl (2009) and Woodward (2003)). The Causal Markov Condition and Faithfulness are presented, and particular attention is drawn to the discovery of casual structure. The manner in which we use classical interventionist methods to determine causal structure and the interpretational stance required to accept the validity of these methods is discussed. This will lead the reader to a deeper understanding of the difficulties interventionist causal theorists face in light of quantum correlations (the subject of Chapter 2), and also an appreciation of why the quantum models of Chapters 3, 4 and 5 are structured the way they are.

Finally, some of the well-known problems with Interventionist causation are reviewed. The purpose of this section is to allow us to later see which problems go through unaltered in the quantum case, and which take on a new character.

1.1 Introduction.

In virtue of *what*, exactly, is a causal model a representation of causal structure? For many interventionists this is primarily a pragmatic matter. Causal models enable one to identify effective strategies by distinguishing between probabilistic correlations that are due to causes and those that are merely accidental. Nancy Cartwright (1979) was arguably the first to clearly articulate the importance of this distinction in her paper “Causal Laws and Effective Strategies”.¹ She was responding to a version of causal eliminativism due to Russell, who had

¹Although see Gasking (1955), Collingwood (1940) and von Wright (1971) for specifically agent-dependent fore-runners to manipulability accounts of causation.

argued that causal principles could not be derived from symmetric laws of association, and thus ought to be eradicated from science.² Cartwright argued against the eliminativist element of Russell’s argument: for her, “causal laws cannot be done away with, for they are needed to ground the distinction between effective strategies and ineffective ones” (p420). In the last thirty years this basic idea has developed into a sophisticated account of causation known as Interventionism. The representational tool of choice for the computer scientists, statisticians and philosophers who utilise this theory is the *causal model*, a graphical structure that has found application across a wide range of disciplines.

For the interventionist, causal models are vehicles for learning about the manipulable elements of the world. It is therefore relevant to ask how one ought to think of the relations and relata of these models. Are they to be considered merely as agent-dependent projections (Price, 2013)), inherently perspectival (Price, 2007) and ontologically deflationary (Price and Menzies, 1993)? Or is it possible to maintain an objective, agent-independent account (Woodward (2003), (2007)) and suggest that causal models may have some (however thin) ontological significance (Woodward, 2015)? While disparate positions have recently been brought somewhat closer (Ismael, 2015), these kinds of metaphysical questions are far from settled. As the primary aim of this thesis is to produce a consistent formalism, one that differentiates between effective and ineffective strategies involving quantum phenomena, I will refrain from engaging in these kinds of questions. Apart from some brief speculation here and there, I will remain agnostic with respect to the metaphysical implications of interventionist causation *per se*.

A number of authors have contributed in important ways to the technical development of interventionism and causal modelling, notably Spirtes et al. (2000), Woodward (2003), and Pearl (2000). We shall spend the remainder of this chapter considering the detail of the latter two accounts. Pearl’s account is mathematically precise and provides a clear formalism to explore: an understanding of this framework will help the reader appreciate the technical issues when using this methodology to characterise quantum causal structure (Chapter 2). Moreover, Pearl’s account is also the departure point for the recent physics literature on quantum causal modelling. As such, it will be needed when we look at these alternative approaches in Chapter 4.

²For the unlucky few who have not heard it, Russells famous quote on the matter: “The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.” (Russell, 1912).

Woodward’s account, whilst arguably less mathematically rigorous, is sufficiently general to motivate the alternative approach that I advocate in later chapters. Additionally, much of the philosophical effort regarding Interventionist causation has focused on this work.³ Both accounts of interventionist causation will pave the way for a deeper understanding of the framework presented in Chapter 5: an interventionist account that allows for the discovery of quantum causal structure.

The great appeal of interventionist methodology is that it closely mirrors the way in which experimental evidence is obtained in many scientific contexts. An example of the technique is seen in Galileo’s famous experiments with inclined planes. The model variables are identified (the size, weight and initial velocity of the objects, and the length and inclination of the planes), the values of some variables are held fixed and the consequences of varying the values of others are observed.

This experimental methodology was also refined by the work of the statistician Ronald Fisher. In the 1930s, Fisher identified the methodological advantages of using *randomisation* to eliminate the influence of latent common causes (Fisher, 1935).⁴ Fisher assumed that randomisation of a variable value would render it independent of its normal causes. Using randomisation to eliminate various kinds of sample bias has since become a key methodology in experimental science. A well known use of this technique is in large, randomised, double blind, placebo controlled, medical trials.⁵ As we shall see, the abstract idea of severing a model variable from its usual antecedent causes is a pivotal component of interventionist methods (section 1.2.2), and one we shall also use in the quantum case (section 5.5).

Our focus is on the advance which came simultaneously from computer science (Pearl, 2000) and philosophy (Spirtes et al., 2000), via the development of Causal Bayesian Networks. These networks represent causal relations among a set of variables via a graph and a probability distribution that factorises over the graph. The graphs can then be used to facilitate various kinds of causal reasoning: they permit the identification of effective and ineffective strategies.

³I leave out the work by Spirtes et al. (2000) primarily for reasons of space. For the interested reader, there is much to commend engaging with the work of these authors in addition to Pearl and Woodward. Whilst the initial discovery algorithms developed by this team are equivalent to Pearl’s Inductive Causation (IC) and IC* algorithms (due to simultaneous and independent work), there have been important and distinctive extensions in recent years. See, for example, Zhang and Spirtes (2015).

⁴Latent common causes are due to variables not included in the model. Their exclusion may be a result of ignorance, or the fact that they are unmeasurable for some practical reason.

⁵Of course, there are well known limitations to this methodology. Specifically, sample bias is only eliminated in the large sample limit.

We start first with Pearl’s formalism, consider his interpretational stance, and then move on to Woodward’s account. Woodward’s work followed on from the formal advances of Pearl and Spirtes et al. and has strengthened and enriched the connection between these technical devices and the traditional concerns aired in the philosophy of causation literature.

1.2 Pearl’s Causal Bayesian Networks

1.2.1 Formalism

Pearl developed his formalism whilst writing artificial intelligence programs. His aim was to capture the manner in which we learn about the world with only limited actions in a noisy environment. Even for a very simple toy environment, inputting and searching through relevant probabilistic data can be extremely onerous. Imagine one has a table of statistical data pertaining to a set containing n variables. If each variable has k possible values, then to exactly specify a probability distribution over *all* possible combinations of values in the model, one needs $k^n - 1$ parameters. Pearl showed that a Bayesian Network (BN) provides a possibility for representing such a joint distribution in a more compact form, by identifying conditional independencies and dependencies. Furthermore, by linking the notion of probabilistic dependence to one of causal mechanism, he demonstrated that one can also utilise Bayesian Networks to uncover explicitly *causal* structure. Structure that can be employed as both a guide to future decision and action, and provide a picture of what might have been.

The relata of Bayesian Networks are classical random variables X_1, \dots, X_n . It is assumed that each variable can be associated with a range of ‘values’: properties that we can unambiguously reveal by measurement or direct observation. Such variables can be binary and used to represent the occurrence or otherwise of an event, can take on a finite range of values or have values that are continuous.⁶ It is generally assumed that the properties that these values represent are non-contextual (in the sense of quantum contextuality) and exist prior to, and independently of the act of measurement or observation.⁷ Ultimately, these values represent the point of contact between the model and the world.⁸

⁶As is the usual convention, capital letters (X_1, X_2, \dots) will be associated with particular *variables*, and lower case letters with particular variable *values* (x_i, x_j, \dots).

⁷Very roughly, contextuality means that the value of the observed property depends on the way it is observed, or on which other properties are observed together with it. See Kochen and Specker (1967) for an introduction to quantum contextuality.

⁸We shall see in sections 2.3.5, and 5.2 that this very intuitive starting point is deeply problematic in the

The BN is an ordered triple $\langle V, G, Pr \rangle$. Here V is the set of variables, G a directed graph and Pr a joint probability distribution over the variables V . The graph captures the qualitative relationships between the variables, with the nodes of the graph being the variables in V and the arrows between them representing the probabilistic dependencies. Some basic terminology will be useful: a variable A is a ‘parent’ of B when there is a single arrow from A to B . In such a situation B is a ‘child’ of A . A is an ‘ancestor’ of B when there is a ‘directed path’ of one or more linked arrows from A to B , in such a case B is a ‘descendant’ of A . The relation of parent to child node, characterised by an arrow, is assumed to represent probabilistic dependence.

The probabilities of BN adhere to the axioms of Kolmogorov’s probability theory (1950): positivity, normalisation and additivity. Intuitively, no event can have a negative probability (positivity), the probability that some possibility or other will occur is 1 (normalisation) and, if two events are independent, then the chances of one do not effect the other (additivity). BN also require conditional probabilities, for example the conditional probability of X given Y , denoted $X|Y$, is the probability of X given the assumption that Y occurs. One can either take such conditionals as primitive (Hájek, 2003) or as a result of the ratio formulation $P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$, which requires that the probability $P(Y)$ be well defined. Two variables X and Y are conditionally independent given a third Z , if the joint conditional factorises: $P(X, Y|Z) = P(X|Z)P(Y|Z)$.

More carefully, a BN is a directed acyclic graph (DAG), G , where each node X_i has an associated conditional probability distribution. This conditional distribution denotes the probabilistic dependence of the values of X_i on its parents in the graph, $P(X_i|Pa_G(X_i))$. Parentless nodes are simply conditioned on the empty set and thus associated with a marginal distribution. A BN represents a joint probability distribution, taken over the model variables, via the product rule $P(X_1, \dots, X_n) = \prod_i P(X_i|Pa_G(X_i))$. In words, the joint distribution (P), taken over all the variables in the model (V), decomposes over the graph (G) into a collection of smaller conditional probability distributions.⁹ This factorisation enables the formal manipulations of the graph that make various causal and statistical inferences possible.

Pearl’s causal models are essentially such Bayesian Networks with some additional assumptions, known as *Acyclicity*, *the Causal Markov Condition* and *Faithfulness*.¹⁰ One requires these

quantum case.

⁹For *Causal Bayesian Networks* these conditional distributions are sometimes known as the *Causal Parameters* of the graph.

¹⁰Faithfulness in Pearl’s terminology is *stability*. I use “Faithfulness” as it is the predominant term in the

assumptions because for probabilistic BNs, directed edges need *not* convey causal meaning. For Pearl “behind every causal conclusion there must lie some causal assumption that is not discernible from the [joint probability] distribution.” (Pearl, 2000)[p332]. That is, without some causal semantics, there is no reason to assume that the purely probabilistic information given to us by Bayesian networks should support causal reasoning. Inferring *causal* dependence from statistical dependence, and *causal* independence from statistical independence requires the further assumption that statistical correlations are generated by local causal mechanisms. Moreover, that we can validate causal structure via localised interventions.

For *causal* Bayesian networks, it is assumed that the joint probability distribution taken over the variables of the system is generated by its ‘casual structure’, where this structure is formed by deterministic causal mechanisms acting between variables, plus some added noise. Such causal structure is represented via the DAG, which is assumed to be isomorphic to the network of autonomous causal mechanisms. When one *passively* observes the system (collects data samples without setting variables to particular values), one is given a window into certain aspects of this structure; when one *intervenes* on the system (some of the data instances correspond to cases where particular variables in the network are set to specific values) typically one gains further information about the structure. It is the latter possibility of causal discovery via local interventions that ties the causal modelling framework to Cartwright’s original insight. The ultimate goal of such models is to give the user a handle on the manipulable elements of the world: to provide a guide to future action. These ideas are made more precise via the Causal Bayesian Networks (CBN) formalism.

What is an intervention?

Interventionist information is obtained when some of the data instances correspond to cases where particular variables in the network are set to specific values, while others are allowed to vary freely. The act of “fixing” variable values must be due to a parameter outside the variable space of the model. The presumption is that the resulting values of the observed variables are determined by the causal structure of the network.

The conceptual distinction between interventions and mere observations is crucial to Pearl’s account. Whilst both can, in principle, be represented as variables, associated with a distri-

philosophical literature.

bution over possible values, interventions must be outside the variable space of the model. Of course, one can always expand the variable set and include these intervention variables explicitly in the model, but to make any causal inferences on this new structure, one needs a further set of intervention variables to test the causal structure of the expanded model.¹¹ Thus, the distinction is relative to where one draws the boundaries of the model under consideration.

Although Pearl characterises the act of intervention as ‘clamping’ to a single variable value, in many experimental situations the best one can do is characterise a distribution over values. Generally speaking, the more sharply peaked this distribution is, the more valuable the act of intervention will be for determining causal structure. Pearl discusses these more general kinds of interventions in Chapter 3, section 3.2.2. In many experimental contexts, rather than forcing a variable X to take on a specific value x_i , the best one can do is induce the value of X to respond in a specified manner to some set Z of other variables. For example, this may be via a stochastic relationship where X is set to x with probability $P(x|z)$, or via a functional relationship $x = f(z)$. We shall later see that these more generalised interventions are needed in the quantum case (see section 5.5).

Acyclicity, The Causal Markov Condition and Faithfulness

Pearl’s chief aim was to identify situations in which one could answer intervention queries using only observational data. For example, imagine one has gathered observational statistics relating four binary variables: male/female, smoker/not, educated/not, wealthy/not. Consider the task of characterising the causal relationships between these four variables, in the case where one lacks the means to directly intervene and force members to adopt particular variable values. An example of an interventional query in this context might be: would the distribution over the wealth variable change if one were to intervene and *force* all the male members of the population to smoke? Without assuming Acyclicity, Faithfulness and the Causal Markov Condition, it is usually not possible to answer these kinds of questions from observational data alone.

As an easy example to illustrate this latter point, consider fig 1.1. The three graphs in figure 1.1 all capture the probabilistic relationship ‘ X is conditionally independent of Z , given

¹¹Obviously, this has some implications for extension to the point where one encapsulates the entire universe. For some, the fact that there is no-longer an “outside” implies that at this level causation disappears (Woodward (2007), Hitchcock (2007)). For others, this is not considered a problem (Frisch, 2014b). The quantum version of this dilemma is discussed briefly in section 5.9.1.

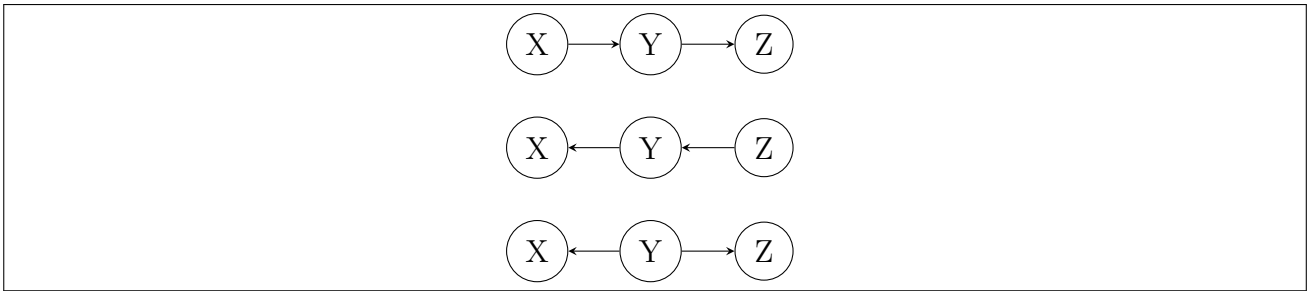


Figure 1.1: Possibilities for a three node DAG assuming causal sufficiency (no latent nodes).

$Y \perp (X \perp Z | Y)$.¹²

So, if we start with purely observational data, and identify the statistical relationship $X \perp Z | Y$, we can only construct an equivalence set of graphs. Such graphs are said to be “independence equivalent”, I-equivalent or “Markov equivalent”. To differentiate between such graphs we need to feed in interventionist information.¹³ In the absence of such interventionist information, for more complex graphs we can also reduce causal underdetermination via the three assumptions of acyclicity, the CMC and Faithfulness. As I make clear below, in Pearl’s view these distinct approaches are actually two sides of the same coin.

Acyclicity is relatively intuitive, simply demanding that no path through the graph is closed to form a loop. That is, causes cannot be their own effects.¹⁴ The Causal Markov Condition (CMC) is also fairly intuitive. The condition states that for a graph G , each variable X_i is causally independent of all its non-descendants, given its parents $Pa(X_i)$ in G . One can think of it as a generalisation of Reichenbach’s screening off criteria (Reichenbach, 1956).¹⁵ That is, it allows the inference from connectedness in the graph (statistical dependence) to causal dependence. We shall consider the CMC and the relationship to interventions in more detail in section 1.2.2 below.

Faithfulness (alternatively *stability* or *no fine-tuning*) states that the only conditional independencies in the distribution P are the independencies that hold for *any* set of causal parameters D . Another way to put this is that all the independence relations in the probability

¹²Symbolically, I use “ \perp ” to denote probabilistic independence (in this case, conditional independence) for the remainder of the thesis.

¹³Using temporal information alone will in general not enable one to distinguish cause from correlation: the mercury falls before it snows, but the falling mercury does not *cause* the snow to fall. Hence, the focus is on interventions rather than temporal information.

¹⁴Although *typically* acyclic, cyclic versions of Bayesian Networks exist (Koller and Friedman, 2009), and can be put to work to determine causal relationships. We discuss these later in section 1.3.5.

¹⁵Figure 1.1 provides a simple example of probabilistic screening off. In each of the three graphs, Y screens off the dependence between X and Z .

distribution over the variables in V must be a consequence of the Markov condition. The idea here is that one does not wish to allow for “accidental” independencies that are created when causal paths cancel. Faithfulness therefore licences the inference from unconnectedness in the graph to causal independence.

To some extent Faithfulness is reflective of model choice in normal scientific practice. For Pearl, causal models should be “chosen by a criterion that challenges their stability against changing conditions, as this is indeed what scientists attempt to accomplish through controlled experimentation” (Pearl, 2009) [p61].

In terms of *discovering* causal structure from sets of empirical data there is another good reason for disallowing unfaithful, or fine-tuned, models. Otherwise it becomes possible to trivially associate any probability distribution with a Markovian graph by ensuring the graph is *complete*, where each node in the graph is connected to all other nodes in the graph. In such a situation, discovery of causal structure becomes impossible.

Recall, causal discovery was one of Pearl's key motivations. In his framework, identification of causal structure can proceed in two equivalent ways: one can either take observational data and assume the CMC and faithfulness, or use data derived via local interventions. Both methods provide the possibility of determining causal structure from empirical data.

1.2.2 Interpretation

Pearl suggests two physical facts underpin his claim that BNs that are Markovian and Faithful can be interpreted to represent causal structure. Firstly, he assumes the local conditional probability distributions (between parent nodes and their respective child node) encode information about local (in time and space) stochastic *mechanisms*, where the stochasticity is due to our ignorance about other (background) variables that we do not include in the model.¹⁶ This means that the relationships between the values of parent and child nodes can be expressed as a function $X_i = f_i(paX_i, U_i)$, where the U_i represent all the unmodelled influences on the variable values (noise). Pearl assumes that if all causal influences are correctly represented, then the U_i will be independent, thus both the Causal Markov Condition will hold and the model will be causally sufficient.

The functional relationship corresponding to a mechanism is taken by Pearl to be *determin-*

¹⁶This will clearly be problematic in the quantum case.

istic: $X_i = f(paX_i)$, where the unknown effect of background variables can be summarised as a single variable associated to each node, U_i . This defines the conditional probability associated with the variable X_i as:

$$P(x_i|pa_{x_i}) = \sum_{u_i} P(x_i|pa_{x_i}, u_i)P(u_i) \quad (1.1)$$

On this interpretation the model is probabilistic by virtue of our ignorance about the values of the U_i : the distribution over the U_i generates the probability distribution over the model.

Pearl is clear that the deterministic causal relations ought to be thought of as primitive, with probabilities entering by virtue of our ignorance. In the preface to his book “Causality”, he admits to previously taking probabilistic relations as primitive, with causal relations entering simply as a useful way of organising such personal probabilistic information. Now, however:

I take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world. (Pearl, 2000)[xvi]

This view relates to the intuition underlying the second physical fact that Pearl commits to: the possibility of localised interventions. He assumes that the act of setting a variable to a particular value (an intervention) can deterministically override the causal mechanisms of the model. Thus, an intervention provides new information by disrupting *only* the local mechanisms associated with that node. It is therefore an influence “that originates from outside the probability space” of the model (Pearl, 2001)[36]. The intervention is assumed to *replace* the original influencing causal mechanisms (incoming arrows) with a mechanism that determines the child variable value $X = x$ with probability 1. The graphical consequence is that arrows into the intervened variable are broken, and a new probability distribution is associated with the altered graph. Such an intervention is called an ‘arrow-breaking’ or surgical’ intervention.

Remember, the joint distribution of the undisturbed graph is:

$$P(X_1, ..X_n) = \prod_i P(X_i|Pa_G(X_i)) \quad (1.2)$$

where the product on the right is over all the nodes in the graph. The joint distribution for an

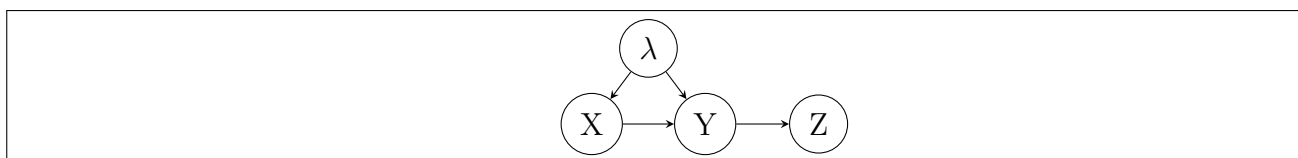


Figure 1.2: Unmutilated graph depicting structure associated with observational information only.

altered graph, for example when X_3 is set to x , is:

$$P(X_1, \dots, X_n | do(X_3 = x)) = \frac{p(X_1, \dots, X_n)}{p(X_3 | Pa(X_3))} \quad (1.3)$$

The term on the left is to be read as “the joint probability over all the variables X_1 to X_n , given that we set the value of X_3 to x ”. Note, it is the factorisation of the joint probability into local conditional probabilities between parent and child nodes that ensures this manipulation of the graph accurately represents the effect of a localised intervention. This licences Pearl’s assumption that “the change is local and does not spread over to mechanisms other than those specified” (p23).

Pearl represents a particular intervention by a variable I added to the graph G under consideration. This variable connects with only one other variable X_i via a directed edge and can take values in $\{idle, do(x_i)\}$, where x_i ranges across all the values of X_i . As a result of setting the intervention variable to a particular value, any incoming casual arrows are removed, resulting in a manipulated graph (see figures 1.2 and 1.3). Such graphs are, rather violently, called *mutilated* DAGs.

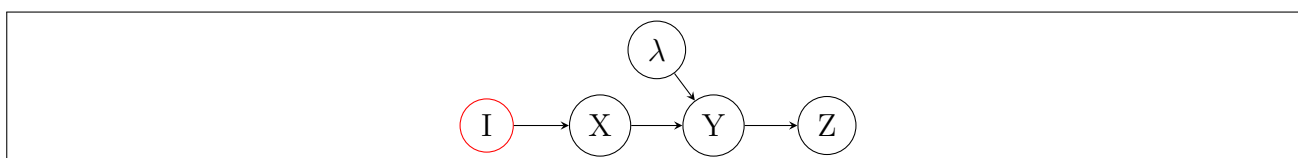


Figure 1.3: Mutilated graph representing the effect of an intervention.

The resulting distribution over the mutilated graph (Figure 1.3) is calculated by assigning the following values to the conditional distribution associated with the node receiving the intervention:

$$P(X_i = x_i | pa_i, I_i) = \begin{cases} P(X_i = x_i | pa_i) & \text{if } I_i = \text{idle} \\ 0 & \text{if } I_i = do(x'_i) \text{ and } (x_i \neq x'_i) \\ 1 & \text{if } I_i = do(x'_i) \text{ and } (x_i = x'_i) \end{cases} \quad (1.4)$$

Pearl formulated the so-called “do-calculus” to formalise the effect various interventions of this kind can have on the probability distribution over the graph. By using this calculus one can make assessments regarding the identifiability of the causal graphs from probabilistic data.

Pearl claims his framework rests on a hierarchy of assumptions, statistical then interventionist. The basic axioms of probability theory (positivity, normalisation and additivity) and the causal assumptions of Markovianity and Faithfulness provide the connection to testing the model via local interventions.¹⁷

The importance of localised interventions in characterising causal relations is usually left implicit in characterisations of the CMC (see equation 1.2). In actual fact, the *causal* information of a DAG is encoded in the conditional probability:

$$P(x_1, \dots, x_n | i_1, \dots, i_N) = \prod_{j=1}^n P(x_j | pa_j, i_j), \quad (1.5)$$

which reduces to equation 1.2 when $I_j = \text{idle}$ for all j .

It is important to realise that the interventions need not be under agent control. So called “natural interventions” are events that are themselves due to causes from outside the variable space of the model. What matters is that the *effect* of such a natural intervention is to sever the target variable from its usual antecedent causes. In his own words:

Nature is a society of mechanisms that relentlessly sense the values of some variables and determines the values of others, it does not wait for the human manipulator before activating those mechanisms. (Pearl, 2009) [p361]

So, for Pearl, the causal structure of a DAG is related to the existence of deterministic causal mechanisms, where the action of a local intervention is to choose between possible individual

¹⁷One can argue whether the ability to perform interventions needs to be in practice, rather than merely in principle. Many authors would consider the latter to be enough.

causal mechanisms.

1.2.3 Causal Mechanisms

Whilst Pearl's commitment to causal mechanisms is clearly important to his overall interpretational stance, the notion of mechanism is left relatively vague:

The world consists of a huge number of autonomous and invariant linkages or mechanisms, each corresponding to a physical process that constrains the behaviour of a relatively small number of variables. Actions are local in the space of mechanisms. (Pearl, 2009)[p223]

Formally, a mechanism is represented by a function relating child to parent nodes. The critical feature however, is the relative *autonomy* of each individual mechanism with respect to others in the network. Recall, mathematically, this is ensured by the factorisation of the joint probability over all the nodes in the graph, into collections of conditional probabilities, one for each node. Pearl does not try to tell us *why* the world should decompose in such a manner, but it is clear that he thinks that this is a primitive ontological fact.

An appealing feature of the mathematical structure of Pearl's algorithms is that they provide a framework that is fairly flexible with respect to interpretation. For instance, the notion of mechanism can be cashed out in variety of ways, all of which remain consistent with the formalism.

Mechanism is a hot topic in contemporary philosophy of science. Andersen (2014) has produced a useful taxonomy of various notions of mechanism. Of the five types, it is not entirely clear which Pearl himself would subscribe to, although it is possible to show that all five remain compatible with his formalism. Andersen initially distinguishes two distinct kinds of mechanism, those that are explicitly anti-reductive (Mechanisms₁ and Mechanisms₂), and those that are not (Mechanisms_{3,4,5}). Anti-reductive mechanisms submit that one can divide the world into different levels, and accept the connections between these levels, without needing to reduce the higher levels or regard them as merely epiphenomenal. [p274]. The view for Mechanisms₁ is typically metaphysically agnostic, focused more on what science *does* than the

way the world *is*. Mechanisms₂, on the other hand, provide an ontological account of what there is, where each mechanism involves causal interactions between entities within it. By looking more closely at these causal relationships we can find generating sub-mechanisms: it's "...mechanisms all the way down." [p280].

The anti-reductionist picture of Mechanisms₁ is perhaps implied by Pearl's assumption of the Causal Markov Condition. Pearl suggests the ubiquity of the Causal Markov Assumption in scientific discourse is "reflective of the granularity of the models we deem useful for understanding Nature" [p44]. Causal models are relative to a level of detail. The Markov condition picks out a particular level at which robust causal relations can be found, in particular, a level at which one can average over unmodelled causes. The mechanisms that define them, likewise, are relative to a specific level of granularity. Woodward also recognises this 'grain specific' feature of interventionist models, and we shall revisit it again in subsections 1.2.7 and 1.3.1 below.

Mechanisms_{3,4,5} are not committed to such an anti-reductionist picture. Mechanisms₃ is defined in Andersen's taxonomy as corresponding directly to Pearl's (and Woodward's) view. The defining feature is modularity of mathematical representation and manipulability independent from other mechanisms₃ in the system. Although Andersen specifically associates Pearl's view with this category, we have noted that it seems plausible that the mechanisms of Pearl's account could also fall into the first, and possibly the second category.

Mechanisms₄ differs from Mechanisms₃ by virtue of the metaphysical picture of causation with which it is typically associated. For Mechanisms₃ causation is difference-making (a cause makes a difference to whether the effect obtains) and for Mechanisms₄ it is productive or physical.¹⁸ As such, the latter view aligns with Dowe's account of causation, where causal processes involve the transmission of conserved quantities such as energy or charge, and causal interactions are exchanges of conserved quantities between two or more causal processes. Mechanisms₄ are simply the conjunction of causal processes and causal interactions [p 288].

Pearl's quote on page 22 above, where he suggests mechanisms correspond to a physical

¹⁸See Ney (2009) and Glynn (2013): for good reviews of both difference-making and process causation, and also possible relationships between these account.

process, implies that he may well view Mechanisms₃ and Mechanisms₄ as equally plausible characterisations. Typically, philosophers see these two accounts of causation as competing rather than complimentary. It is clear, though, that Pearl's account of causation can, to some extent, fit with both.

We shall see in later chapters that the quantum causal modelling framework presented provides a similarly metaphysically flexible notion of mechanism.

1.2.4 Type vs Token causation

Should one view Pearl's theory as pertaining to *token* or *type* causation? Tokens are particular, unrepeatable events. To take a well-known example, the event of Socrates death by drinking hemlock. Types are events that are repeatable: death by drinking hemlock. Token-causal claims relate event tokens and type-causal claims relate event types. Typically, philosophers take these two kind of causation (also known as *general* and *singular* causation) as belonging to distinct species. Pearl however, sees the distinction as a matter of degree (Pearl, 2009)[p301-311]. For example, if one has full knowledge of the environment (the values of each of the U_i relative to the instance of the singular causal claim under consideration) then one can make statements pertaining to singular causation in full confidence. Given partial knowledge of such background factors, one is consigned to probabilistic statements, and the associated causal claims become generalised. The degree is therefore relative to how much relevant background knowledge we possess for the given example case.

1.2.5 Discovery Algorithms

Recall, Pearl's aim was to identify situations in which one could answer intervention queries using only observational data. That is, he wished to identify causal structures that could act as guides to action, and could enable the user to distinguish between effective and ineffective strategies. He developed two well known algorithms that take as input a list of conditional independencies (found in a joint distribution over a given set of variables) and return a set of DAG's as output. The Inductive Causation algorithm (IC) will return a DAG, under the

assumptions of Causal Markovianity, Faithfulness and Causal Sufficiency (no unmeasured common causes). The IC* algorithm does not require the assumption of causal sufficiency, but will in general only return a partial ancestral graph (PAG): a DAG with any number of undirected edges. These algorithms use observational data only, so any remaining underdetermination of the causal structure can be further reduced if one has access to interventionist data.¹⁹

The greater level of casual underdetermination resulting from the IC* algorithm is due to the possibility of unmeasured common causes, or *latent variables*. Latent variables considerably complicate the task of discovering causal structure. As a simple illustrative example, consider allowing for a single latent variable. Further possible Markov equivalent graph structures that capture $X \perp Z | Y$ are depicted in figure 1.4.²⁰

The increased number of graphs highlights the complexity cost of using conditional independence relations found in observational data to recover causal structure when one allows for latent variables. Typically one does not recover a unique DAG when allowing for such a possibility.

1.2.6 Dependence separation

There is one final aspect of Pearl's formalism to review: so-called *d-separation*. We shall need this in Chapter 4, when we look at some proposals for quantum causal models that have recently appeared in the physics literature. D-separation (d for *directional*) provides a useful tool for checking which variables are independent and dependent (both marginally and conditionally) for complex graphs.

As stated above, Pearl's algorithms are able to construct causal graphs from statistical data sets via searching for probabilistic dependencies and independencies between various sets of variables. The algorithms proceed in a recursive manner, and search first for variables that are probabilistically dependent, conditioned on the null set. This allows the generation of a graphical *skeleton*; a graph with undirected edges. Next, the algorithms search for *v-structures*, triples of the form $x \rightarrow y \leftarrow z$. Such structures are also sometimes known as *unshielded colliders*

¹⁹Although see Eberhardt (2013) for some interesting problem cases.

²⁰Eberhardt (2007) p19

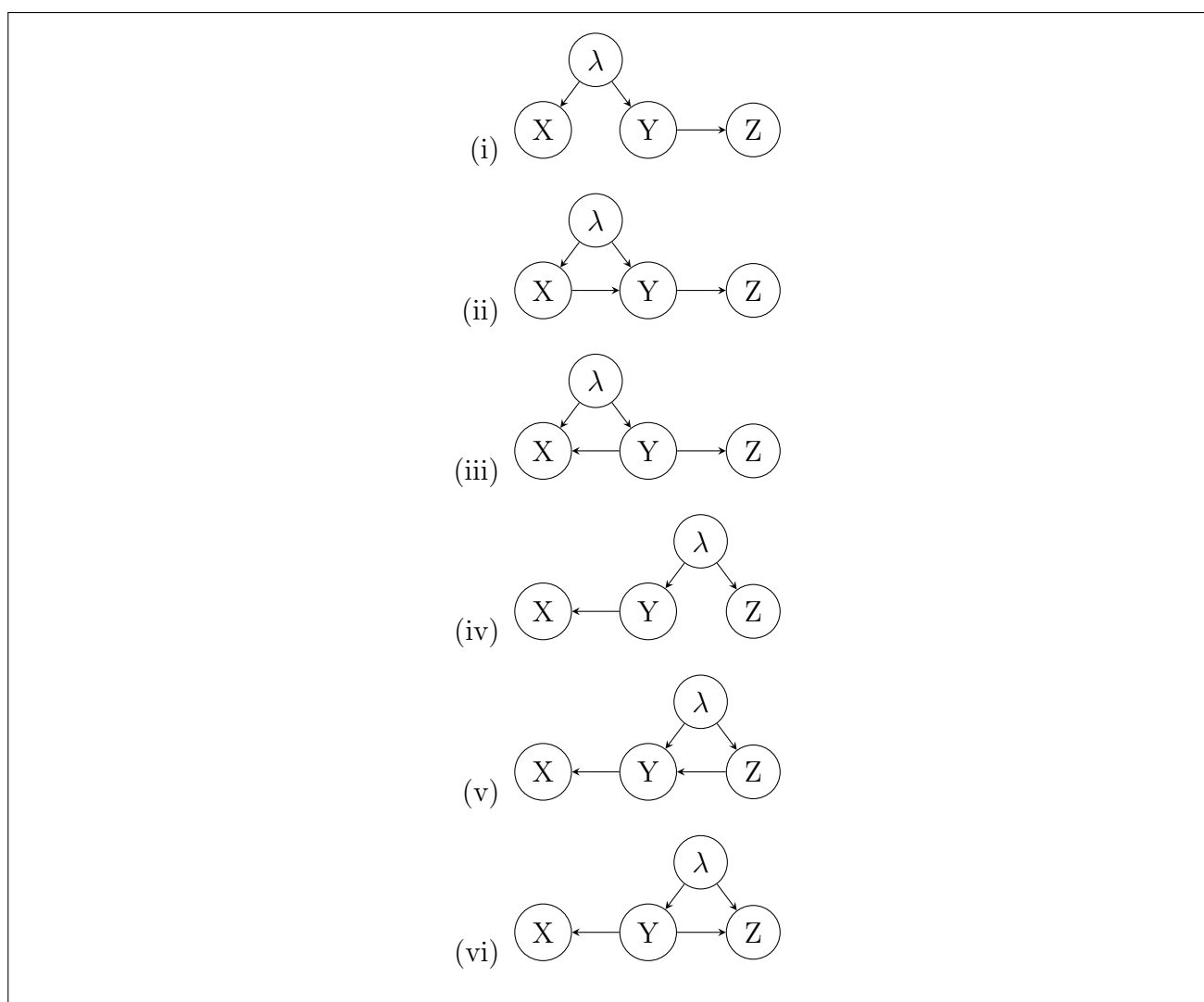


Figure 1.4: Possible three node DAGs when one allows for latent nodes

to emphasise the importance of the missing arrow from x to z . The reason why the algorithm proceeds in this manner is intimately related to the criterion of d-separation.

It is possible to get an intuitive feel for how d-separation works by considering possible probabilistic relationships between the three node graphs in Figure 1.5. The goal of d-separation is to determine when two sets of variables are independent, given a third set. Searching for such sets forms an important step in constructing causal DAGs from probabilistic data. Consider the four possible arrangements of three variables (figure 1.5).

For the first three, X and Z are marginally dependent, but independent given Y . For the fourth, a structure called a collider, X and Z are marginally *independent*, but conditionally dependent given Y . D-separation exploits this asymmetry and searches for such so-called col-

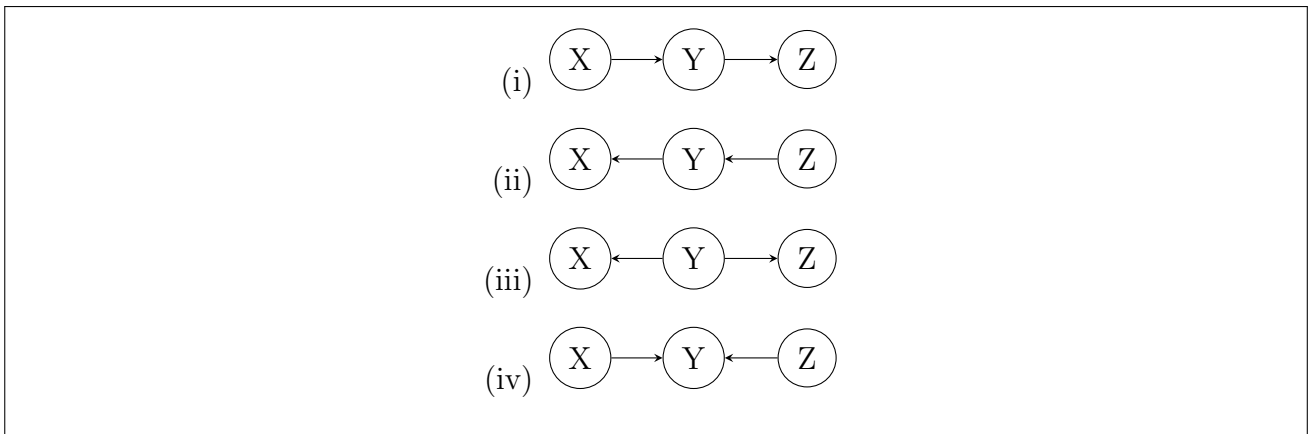


Figure 1.5: Possible DAGs using three nodes. The first three DAGs reflect marginal dependence between X and Z , but independence conditional on Y . The third DAG depicts marginal *independence* between X and Z , with *dependence* given Y .

luder structures; if the conditioning set contains such collider variables then d-separation fails. This criterion can be extended to apply to sets of variables. In this manner one can discover independencies further to those that may appear initially obvious for any given graph structure.

The value of this criterion is two-fold, if a probability distribution P decomposes over a graph structure G in accordance with this criterion, then the graph will obey the Causal Markov Condition. Additionally, it provides a tractable method for checking causal independencies in complex graph structures, and for verifying when two causal DAGs are observationally equivalent.

1.2.7 The Bigger Picture

Let us take a step back, and look at two general features of Pearl’s formalism. Firstly, what are the points of contact between such causal models and the world? Roughly speaking there are two: the data that underlies the variable ‘values’, and the data that we use to characterise the local interventions. It is worth noting here that neither kinds of data are explicitly included in the final model: it is the axioms of probability theory that get us from the data to the final model.

Secondly, and more importantly (for our purposes at least), is the fact that Pearl has in effect given us an iterative account of scientific modelling. One does not arrive at a causal model in a conceptual vacuum: typically one does not either just “start with the data”, so to speak,

and produce a fully fledged causal model. Nor does one start with a plausible causal structure, according to domain knowledge and theory, and expect it to perfectly match an empirically derived joint distribution. Rather, the process is one of mutual refinement, where model and theory are developed in a co-evolutionary manner. Pearl's techniques capture this feature of causal modelling in a mathematically consistent manner:²¹

The Markov condition guides us in deciding when a set of parents $Pa(X_i)$ is considered complete, in the sense that it includes ALL the relevant immediate causes of X_i . It permits us to leave some of the causes out of $Pa(X_i)$ (to be summarised as probabilities), but not if they also influence other variables modelled in the system. If a set $Pa(X_i)$ is too narrow, there will be disturbance terms that influence several variables simultaneously, and the Markov property will be lost. Such disturbances will be represented as latent variables. Once we acknowledge the existence of latent variables and represent their existence explicitly as nodes in a graph, the Markov property will be restored. (Pearl, 2009)[p44]

There are several avenues for restoring the Markov property to an otherwise recalcitrant causal model: one can fine-grain variable values, increase the scope of the model or look for correlated noise. If our empirically derived data sets are not giving us Markovian graphs, or are giving us graphs that do not correctly support interventions, then we have a variety of avenues to explore in order to achieve a more representational causal model.

The fact that the Pearl's formalism reflects the iterative manner in which scientific models develop can also shed light on a well-known philosophical worry with interventionist causation. Many have argued against Faithfulness, claiming it is too strong. There are many cases of models of physical systems that appear fine-tuned, despite the fact that intuitively we still wish to call them causal (Cartwright and Jones (1991), Eberhardt (2013), Andersen (2014)). There is a subtlety that somewhat mitigates this concern however. As we have seen, causal models are relative to a number of pragmatic choices: the model scope, the range of invariants

²¹This is also the case for the modelling strategies utilised by Scheines, Glymore and Spirtes (2000).

(background variables we exclude from the model) and the level of detail.²² For any fine-tuned model it will in general be possible to recover a faithful (stable) model by either changing scope, increasing the level of detail or altering the range of background invariance (assuming Pearl's interpretation is correct). Thus, in classical modelling situations, the Markov condition and Faithfulness guide us toward the most efficacious level to express the causal structure of a given system and play an indispensable part in the discovery of causal structure.

Recall, a fine-tuned, or non-Markovian model does not imply that we have settled on the only representational strategy for the phenomena in question. Rather than signal a lack of causation, such failures can instead be seen as an indication that more work needs to be done. Searching for hidden variables, fine-graining measurement values, analysing noise models and improving intervention control are all viable options for arriving at an improved model. When the set of causal variables is expanded, due to the inclusion of previously hidden (latent) variables, new causal relations can become apparent. Similarly, fine-graining measurement values will change the sensitivity of the model: weaker causal relations will play a more significant role and again new causal relations can appear. Furthermore, when the scope and level of detail of a causal model are altered, error variables may become correlated and more specific noise models are needed to uncover hitherto unnoticed spurious causes. Under such circumstances, once again, the causal relations of the model can change. Finally, by improving the sensitivity of the intervention control, one can sometimes discriminate between alternative models. Fine-graining variable *values* gives rise to plausible alternative models, but we can only verify such structures when we possess the appropriate level of intervention control. Ideally, one needs to be able to intervene to *set* the variable values precisely enough for such fine-graining to count.

In each case, in the face of such refinements, the discovery of alternative causal structures gives more resources to examine the phenomena in question. This process enriches both our knowledge of the relevant domain, and also our knowledge of how best to probe it. It gives us the ability to distinguish between effective and ineffective strategies for a variety of different situations, where the differences are simply due to scope, precision or detail, and reflect our

²²Invariants is Ismael's term (2015). The range of invariants is simply the range of values that the background variables U can assume such that the causal relations of the model still hold. Woodward uses the term *stability* for this concept (see section 1.3.1).

particular concerns or technological capabilities. Valuable work is being done in the field of causal modelling to improve and extend the reach of such techniques.²³

To understand the strength of Pearl's formalism, it is important to recognise that the alternative models produced by changes in scope or detail need not be in conflict with one another. Despite the superficial appearances of changing causal relations, these models can be seen as reflecting complementary representations of the same objective reality. The justification of this claim rests in the fact that each model is determined by a movable boundary, a boundary defined along two dimensions by (i) those variables that are included in the model and those that are left out, and (ii) the level of detail. Typically, this choice is made on pragmatic grounds. Whether or not this necessarily renders the *objectivity* of causal models hostage to agential concerns is a difficult issue, and although intriguing, not one I shall engage with in this thesis.²⁴

1.2.8 Summary

In this section I have attempted to capture the key components of a very rich and complex body of work. Pearl has published hundreds of papers and 3 books on the topic of causation and causal modelling. As such, I have really only presented the briefest of overviews. The details I have presented here are simply those that I feel link most closely with the issues associated with producing quantum causal models. The use of the CMC and faithfulness to uncover causal relations from observational data is a key feature of Pearl's methods. The link to local interventions is made via postulating the existence of autonomous causal mechanisms. Such mechanisms are assumed to be deterministic, with probabilities entering by virtue of our ignorance of all the facts.

We now turn to an alternative approach to interventionist causation: that of the philosopher James Woodward.

²³See, for example, Zhang and Spirtes (2015).

²⁴Although, see (Ismael, 2015) for an interesting perspective on this debate.

1.3 Woodward's causal models

1.3.1 Methodology

Woodward's approach to causation is closely aligned with that of Pearl. His highly influential book "Making Things Happen" was motivated by a perceived lack of clarity and consensus within the philosophical field of causation and causal-explanation (Woodward, 2003).

What Woodward saw was an opportunity to bring the philosophy of causation into closer contact with important work that was being done in other fields, whilst simultaneously solving some of the problems that had beleaguered traditional approaches. The downside of having such broad goals, of course, is that one typically ends up presenting many avenues for attack. Despite over a decade of critical review, however, Woodward's account still finds favour with many philosophers of science as the preferred approach to causal enquiry.

Very roughly, his theory trades on the intuition that we typically associate causal relations with an ability to manipulate or control the value of a particular variable. In Woodward's view, causal structure captures information about hypothetical interventions. As such, there is also a very natural fit between almost all experimental contexts and Woodward's theory. To identify a causal relation between two variables A and B , given some particular background conditions Z , he claims we typically wish to both isolate A and also vary its value. Furthermore, we wish to hold the background conditions Z as fixed as possible to eliminate confounding factors, such that we can observe and measure the direct effect A has on B . In this way, one is able to deduce the causal relation by seeing how B changes when we vary A in some way. Following Pearl, Woodward realised that it is the potential for such manipulations that allows us to differentiate between relationships that merely express correlation and those that involve causation.

Woodward's definition of a direct cause, is tied to the definition of intervention:

A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set V is that there be a possible intervention on X that will change Y (or the probability distribution of Y) when all the other variables in V besides X and Y are held fixed at some value by interventions. (Woodward, 2003)[55]

Clearly, it is critical to his theory that one has a formal account of what constitutes an ideal, hypothetical intervention.

The ideal intervention (I) must meet the following requirements (Woodward, 2007)[75]:

1. I must be the only cause of A - the intervention must completely disrupt the causal relationship between A and its previous causes, so that the value of A is set entirely by I .
2. I must not directly cause B via another route
3. I should not itself be caused by any cause that affects B via a route that does not go through A
4. I leaves the values taken by any causes of B unchanged, except those that are on the directed path from I to A to B , (should this exist).

Requirement 1 ensures the value of A is not effected by any other causal variable Z , but entirely set by I . Requirements 2 and 3 remove any other ways A and B may be correlated other than A 's causing B . The causal effect here is relativised to a background context Z (which will incorporate information about other causes of B). Thus this account attempts to answer the question “in a given context, what is the difference made to B by varying A ?”.

The relata in Woodward's causal claims are called ‘variables’, where a variable is simply a property or quantity capable of taking two or more ‘values’. When pressed to clarify the philosophical nature of the variables, Woodward falls back on his characteristically pragmatic stance:

The problem of variable choice should be approached within a means/ends framework: cognitive enquiries can have various goals or ends and one can justify or rationalise candidate criteria for variable choice by showing that they are effective means to these ends. (Woodward, 2015a)[5]

This kind of “means/ends” approach is iterated by Woodward in a number of contexts, and has become a popular approach to various problems in the philosophy of causation.²⁵ Broadly

²⁵See, for example, Hitchcock (2012) and Woodward (2015).

speaking, this is also how we identify quantum causal variables and values in the formalism of chapter 5: by virtue of their pragmatic value.

Woodward quite rightly points out that typically we wish to know more about our causal models. We don't only wish to know which variables are causally related, but "... we wish to know which kind of interventions on X have an effect on Y , how they effect Y and under what background circumstances." (Woodward, 2007)[p76]. This information is captured by two further requirements: invariance and stability. A causal model is *invariant* if it holds under at least some range of values of the intervention. A model is *stable* if it holds for at least some range of changes in background variables, where these variables are typically not included in the model. These requirements capture the grain-sensitivity of causal models that we saw present in Pearl's causal models. The range and level of precision of the interventions, plus the range and level of detail of the background variables ("error variables") ultimately determine the casual relations between variables in the model.

So what of the Causal Markov Condition and Faithfulness? We saw earlier that Pearl's methods rely on these assumptions. In earlier papers, Woodward (along with co-author Daniel Hausman) claimed to be able to *derive* the Causal Markov Condition from the definition of an ideal intervention, a modularity condition (MOD*), plus some other assumptions (Hausman and Woodward (1999). Hausman and Woodward (2004b)). Simply defining causal relations with respect to the possibility of an intervention was recognised as not providing an analysis of causation: although sufficient, one lacks the requisite necessary condition. It is the possibility of fine-tuning (although not identified as such) that prevents the contrapositive: "the corresponding necessary condition is false, because the causal influence of X_i on X_j along different paths may cancel out". As an alternative means of ensuring Faithfulness, Hausman and Woodward defined a new condition, MOD*.

[MOD*] says that when X_i does *not* cause X_j , then the probability distribution of X_j is unchanged when there is an intervention with respect to X_i . We shall show that MOD*, some strong assumptions concerning the unrepresented causes, and two widely accepted assumptions concerning probabilistic causation imply [the

Causal Markov Condition]. (Hausman and Woodward, 2004b)[149]

The “strong assumptions ” here essentially refer to the same condition Pearl uses: any error variables (omitted causes, U_i) must be uncorrelated. Such error variables are direct causes of exactly one of the variables in the model and are causally connected with exclusively that variable and its effects. The “two widely accepted assumptions concerning probabilistic causation” are (i) the determination of probabilities assumption: the distribution over each variable is determined by its parents and any causes not included the model (for deterministic causal relations $X_i = f_i(pa(X_i), U_i)$) and for indeterministic relations $P(X_i) = g_i(pa(X_i), U'_i)$), and (ii) no accidental correlations: there is always a causal explanation of correlations. This final assumption is, by their own admission, strictly not necessary as it is already implicit in the assumption of MOD* (Hausman and Woodward, 2004a, p. 852).

The similarity to Pearl's perspective ought to be fairly clear by now. Both Pearl and Woodward justify the Causal Markov Condition by assuming the existence of local functional relationships (Pearl's “mechanisms”) with probabilities entering by virtue of the effect of unknown environmental variables. As Jenann Ismael notes , where there are differences, they are due primarily to a difference in focus:

.... for Pearl, once you know what the causal mechanisms are, you can say which interactions constitute interventions. Woodward ... wants to characterise the notion of an intervention independently so that it can be used as a probe for causal structure. To some extent this in-house dispute reflects a difference in focus. From a metaphysical perspective, it is natural to take the underlying causal structure as basic. It is what explains the surface regularities and patterns of counterfactual dependence. But Woodward is interested in using interventions as a route in, so to speak. He wants to be able to identify interventions (perhaps provisionally) before we have a detailed understanding of the causal structure and use them to probe.

(Ismael, 2015)[252]

1.3.2 *Philosophical Worries*

Philosophers have had a number of concerns about Woodward's Interventionist theory. To address them all in depth would leave no room to discuss quantum mechanics, so I shall just provide an overview. I provide this not in order to defend Woodward's particular brand of interventionism in any detail, but rather in order to show later which concerns simply carry over to the quantum case, and which take on a new character.

1.3.3 *Circularity*

The most obvious criticism of Woodward's approach is that it is circular. Interventions are themselves causes. Thus the theory defines interventions in terms of causes and causes in terms of interventions. Woodward and Hitchcock both defend this charge, claiming that whilst circular, the account is not viciously so. Woodward is clear that his aim is not to produce a reductive account of causation, where causal terms are reduced to acausal ones, but rather to provide an illuminating account of the relationship between scientific practice and causation. In a sense, it seems he is content to utilise both causes and interventions, without the need for any further reduction. He claims there are clear *methodological* motivations for adopting this view, and in more recent work illustrates nicely (in my opinion) the fact that in the context of causation, metaphysical and methodological philosophical concerns are perhaps not so far apart as one might think (Woodward, 2015).

1.3.4 *Causal Markov Assumption*

Several philosophers have had worries related to Woodward's derivation of the Causal Markov Assumption. Steel (2006) concludes that Woodward and Hausman's derivation of the CMC secures only pairwise independence. That is, variables are only *pairwise* independent of their non-descendants, conditional on their parents [p221]. Steel argues that this weaker form of the CMC does not, in fact entail the CMC as it is usually written, and the consequences are serious. If one only demands pairwise independence then the joint distribution will not necessarily factorise into the product of local distributions over each variable conditional on its parents. In

Hausman and Woodward's derivation, they suggest that the independent error idealisation can be derived from their modularity condition, but as Steel points out, this only secures pairwise independence of the error terms. He concludes that the Causal Markov condition thus

... rests on two assumptions: acyclic causal structure and probabilistically independent error terms. Thus, any deeper justification of the CMC than that currently available must either devise a generalised version of the CMC that can be shown to hold for cyclic causal structures, or provide an account of why exogenous error terms should be independent. [229].

I am not too worried by the former point; it is not my aim to discuss situations involving closed time-like curves or other exotica. The fact that causes can not be their own effects is an assumption I am happy to make. The second worry, however, is more interesting. Why *should* error terms be independent? And will this reasoning transfer through to the quantum case? Recall, for Pearl, uncorrelated error variables are an interpretational requirement needed to secure the Causal Markov Condition. For Pearl, without Markovianity one loses the advantages of manipulability. On non-Markovian models Pearl comments:

... such models - even if any exist - would have limited utility as guides to decisions. For example, it is not clear how one would predict the effects of interventions, save for explicitly listing the effect of every conceivable intervention in advance. [p61]

On this perspective, uncorrelated error variables are an *epistemic* requirement, an assumption that is needed in order to reflect our ability to discover causal structure via interventions.²⁶

1.3.5 Faithfulness

Andersen (2013) presents a different worry. She points out that many biological situations involve feedback loops, and such loops create stable equilibrium states that violate Faithfulness.

²⁶There are, of course, useful non-Markovian models in science. Pearl's comments reflect the fact that such models carry less causally useful information than their Markovian counterparts.

Moreover, such violations are typically stable in the face of external perturbation, and consequently immune to the measure-theoretic proof of Spirtes et al. (2000).²⁷ The models she is referring to, however, violate the assumption of acyclicity, and so the violation of Faithfulness is easy to explain. For small enough time slices, one can produce a Markovian, Faithful model of the phenomena in question. Of course, this kind of fine-graining will likely obscure much of the interesting behaviour occurring at the coarse-grained level. Andersen highlights an important point, however: in order to ‘catch’ such apparent violations of Faithfulness via either interventions or more fine-grained measurement, the timescale of the intervention/measurement must match (or be shorter than) the timescales required for re-establishing equilibrium.

The value of capturing such feedback causal loops has not gone unnoticed by the scientific community. Dynamic causal nets capture the possibility of such stable relations and are reasonably popular among biologists (see, for example Rajapakse and Zhou (2007) and Needham and Westhead (2007)). Such models have a version of d-separation, are Markovian, and capture the temporal characteristics of network connectivity (Koller and Friedman, 2009). They specify a prior Bayesian Network that expresses initial conditions, and a transitional Bayesian Network that captures how the variable values change at each successive time-step. The causal relationships for the so-called template models that represent each time step, however, are both Markovian and Faithful.

1.3.6 General objections

Various philosophers have also objected to Woodward’s presentation of causation in more general terms. It has been said that the interventionist account is metaphysically bankrupt and objectionably operationalist (Baumgartner and Drouet (2013), Strevens (2008), Cartwright (2006)). Process causal theorists, who take causation to be essentially about physical processes and interactions, are uneasy about the lack of “biff” quality in Interventionism (Dowe, 1999). Others (Price (2007) Price (2013)) believe it is better to reduce interventionism to a primitive notion of agency: for these authors causes are more about us than Woodward cares to admit.

²⁷This proof shows that the set of unfaithful models are Lebesgue measure 0 with respect to the space of possible causal models.

There is a large and complex literature addressing these various bones of contention. Rather than engage in them here, I shall adopt Woodward's own perspective on the matter: one ought to think of the interventionist account as providing what he chooses to call a 'functional approach' to the philosophy of causation (Woodward, 2012). In a number of places, Woodward shows that understanding the *practical* advantages of using causal reasoning, identifying causal relations and producing causal models according to Interventionist methodology can shed light on a variety of traditional philosophical concerns about causation (Woodward (2012), Woodward (2015), Woodward (2015b)). I see the methodology provided by interventionist causal theory as contributing to the overall observation Nancy Cartwright made in the 1970's: such methods enable us to distinguish between effective and ineffective strategies.

1.3.7 A cautionary note

This thesis should *not* to be read as presenting a knock-down argument against the Russellian perspective. Recall, Russellians believe that characterisation of initial states, plus laws of dynamical evolution give us a full account of fundamental physics. From this perspective, there is no ontological space for such things as causal relations and relata: they should simply be eliminated. For the most part, what I have to say leaves room for both Russellian *and* anti-Russellian views. Russellians can regard the quantum causal models I present in Chapter 5 simply as representational devices that give us nothing more than a handle on the practical advantages of causal reasoning involving quantum systems. Anti-Russellians can take the view that the pragmatic value of such characterisations should not be so easily dismissed, and follow the line of Nancy Cartwright's original objection.

I do, however have two small, and rather speculative, points to make regarding a possible methodological advantage that the interventionist causal approach may have over a straightforward Russellian view. Firstly, regarding *fundamentality*: it is clear that our current best physics does not involve anything *like* a single, unified, fundamental theory, with an unproblematic ontology to match. The difficulties inherent in producing a theory of quantum gravity are well-known. One obvious problem is that general relativity implies the existence of a de-

terministic, dynamical background causal structure, leaving no *obvious* way to incorporate quantum indeterminism or indeed other quantum properties such as superposition. In part, the recent research into characterising quantum causal models has been driven by the thought that perhaps background causal structure is not “fixed” in the manner described by GR, but rather allows for causal structures that are in some sense *indefinite* (Oreshkov et al., 2012). If indeed it turns out that this is the right way to forge a new, unifying, physical theory, there is a clear role for quantum causal models. As it currently stands, quantum theory does not, by itself, give us a method to determine causal order between arbitrary quantum events. For any new fundamental theory, we would like to verify that in the appropriate limit the theory gives back ordinary quantum physics. That is, a theory with the usual determinate causal order and time evolution of quantum states. A quantum causal modelling formalism provides a way to meaningfully ask what the causal relations between quantum events are.

By contrast, ordinary quantum mechanics (under the Russellian picture of states evolving according to global laws) gives us no handle to approach this problem. Indeed some of the conceptual issues in formulating theories of quantum gravity can be traced back to the difficulty of understanding the empirical content of theories that do not have an external time, nor an absolute causal structure. Given such concerns, it becomes difficult to understand how the Russellian approach of “fixing initial states and then evolving them forward” remains meaningful as providing a picture of *fundamental* physics. The characterisation of quantum causal models, by contrast, provides a useful starting point from which to consider such questions.

The second point of note is simply pragmatic. As engineered quantum networks become increasingly complex, the task of characterising the order in which quantum events take place will become increasingly important (and difficult).²⁸ Current tomographic methods do not provide a method for determining such an order (see Chapter 5) and it is not clear, at least to me, how the Russellian perspective provides any purchase on tackling this kind of problem.

²⁸Classical distributed computing currently uses specific protocols to keep track of the order between local operations (e.g. Lamport time-stamps and vector clock methods). We currently have no analogous methods for characterising the order of quantum operations.

1.4 Summary

We have considered interventionist causation from the perspectives of two theories and noted a number of similarities. Both commit to the idea that localised interventions are the best way to confirm causal structure. It is hard to ignore the success of causal modelling according to the above methods. There is hardly a discipline that has escaped its reach. One can find causal models in philosophy, statistics, computer science, the physical sciences, medicine, economics and psychology (Korb and Nicholson (2010), pp 142-143). And this list is far from exhaustive. There is however, an important problem. In their current form they do not work for systems involving quantum phenomena.

Given what we now know about classical causal models, we can suggest some desiderata we might wish a quantum causal modelling formalism to satisfy :

- (i) The formalism should allow for the *discovery* of causal structure from empirical data. At a minimum, such discovery should be possible using interventionist data (data instances where local events are under external control. It would be an advantage if causal structure could also be discovered in situations where interventionist information was incomplete.
- (ii) *All* correlations between empirically derived data should be accounted for via notions of direct, indirect or common cause relations, i.e. there should be no “unexplained” correlations. In situations where all correlations between empirically derived data can *not* be accounted for via direct, indirect or common cause relations, there should exist a method for extending the model to include possible *unobserved* nodes in order to account for the correlations.
- (iii) Classical causal models should be recovered as a limiting case of quantum ones.

Chapter 2

Interventionist Causation and Quantum Mechanics

In the last chapter we saw that causal structure is intimately linked to the possibility of localised intervention. DAGs are thought to represent causal structure just in case they are able to support intervention queries in the right way. In the classical case, if one only has access to observational data, this requirement can be captured via two key assumptions: the Causal Markov Condition and Faithfulness. In this chapter we put this methodology to work in a quantum mechanical context. Following tradition, we shall examine one of the Bell scenarios in order to determine how classical interventionism fails. Noting that the usual response is to demand that one drops either the CMC or Faithfulness, I shall suggest neither approach is particularly satisfactory and argue for an alternative. We also briefly situate the discussion within the context of alternative approaches to understanding Bell's theorem from outside the causal modelling literature.

2.1 Introduction

There is a large literature discussing the various problems quantum phenomena create for theories of causation. As Richard Healey (2009) puts it: “There is widespread agreement that quantum mechanics has something radical to teach us about causation. But opinions differ on

what this is.” [p673]

The many philosophers who have engaged with this question have done so almost exclusively via analysis of the Bell experiments. Such work has shown it is exceedingly difficult to apply classical causal techniques to quantum systems. Space-like separated entanglement correlations seem to defy causal explanation, and it seems near impossible to produce an account that avoids the difficulties posed by non-locality, contextuality, and the measurement problem. In the last thirty years or so, philosophers have advocated a number of fixes, arguing for non-local common causes (Suárez and San Pedro (2011), Egg and Esfeld (2014)), non-screening off common causes (Butterfield, 1992) and “uncommon” common causes (Hofer-Szabó et al. (2013), Näger (2015)). Others have argued for more exotic solutions such as retro-causation (Evans et al. (2013), Evans (2014) and super-determinism (’t Hooft, 2009).

In this chapter, we shall specifically engage with the difficulties quantum phenomena pose for interventionist accounts of causation. We first revise the detail of the Bell experiments and present some standard responses from the relevant literature. We then consider two interventionist responses: those of Glymour (2006) and Wood and Spekkens (2014). We ask whether these responses, namely that one must choose between the CMC or Faithfulness, get to the heart of the issue and suggest that they do not. Finally, we consider a different perspective and advocate a new direction of research (section 2.3.3). We shall follow this line in the remaining chapters.

2.2 The Bell experiments.

Although the physical setup of the Bell experiments are relatively easy to understand, a great deal of controversy surrounds the appropriate way to interpret the empirical results. These experiments were inspired by Bell’s *theorem*, originally presented in his 1964 paper (Bell, 1964), and written in response to Einstein’s 1935 thought experiment (Einstein et al., 1935). It is worth a brief detour back to this early work in order to better situate Bell’s results.

The Einstein, Podolsky and Rosen paper (EPR) considers measurements performed on space-like separated, entangled quantum particles. They show that for certain sets of measure-

ments, quantum mechanics predicts that the outcomes at each wing will be perfectly correlated for each run of the experiment. That is, knowing the outcome on one wing allows one to accurately predict the outcome for a certain measurement on the other wing.

The EPR argument is then presented as a reductio. The following three assumptions, coupled with the results of quantum experiments, lead to a contradiction.

- (i) Locality: space-like separated systems cannot influence each other (“...no real change can take place in the second system in consequence of anything that may be done to the first system.” [779]),
- (ii) Completeness: quantum mechanics is descriptively complete, and
- (iii) Reality: “If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.” [77]

For Einstein et. al., quantum mechanics refers to such elements of reality via the eigenvalue-eigenvector link: an eigenvalue of a quantum system prepared in the relevant eigenstate will result in an outcome that can be predicted with probability one. In the face of the contradiction, in order to save locality and reality EPR advocate we give up the second assumption: completeness.

Bell’s theorem (in its various forms, see Bell (1964), Bell (1971) and Bell (1976)) can be understood as refuting incompleteness as a way out of the contradiction. The standard interpretation of Bell’s various writings then (in the philosophical literature, at least) is that *any* theory that can account for (i.e. *explain*) the empirical predictions of quantum mechanics must be non-local. So, broadly speaking, one faces a choice between giving up *locality* or giving up *explanation*.

Bell wrote many papers regarding the philosophical consequences of quantum theory, and much ink has been spilled over the various philosophical assumptions and implications of his work.¹ The recent fifty year celebration of his famous 1964 paper have inspired another burst

¹For a comprehensive collection of Bell’s papers, see “Speakable and Unspeakable in Quantum mechanics”, Bell (1990).

of academic activity, and this work suggests that it is not entirely straightforward to gauge exactly what Bell *meant* by his various characterisations of locality and local causality. Indeed, intelligent and careful analysis has resulted in (sometimes alarmingly) disparate positions (see, for example, Werner (2014a), Werner (2014b), Maudlin (2014b), Maudlin (2014a), Wiseman (2014), Wiseman and Rieffel (2015), Norsen (2015)). I suspect this reflects, to some extent, that Bell recognised the philosophical difficulties that confound a precise characterisation of both causation and causal explanation, and was honest about his various qualms. Broadly speaking, his work shows that the connections between the various concepts involved (e.g. instantaneous action-at-a-distance, locality, local causality and agency) are far from obvious.

For our purposes, it is relevant to consider some of Bell's comments in light of our characterisation of interventionist causation in Chapter 1. Most pertinent are two of his later papers: "The theory of local beables" and "La Nouvelle Cuisine" (Bell (1976) and Bell (1990)). I take the view that these papers nicely capture the elements of interventionist thinking that often underly our intuitive notion of causation and causal explanation.

Bell's notion of "local causality" (and, indirectly, causal explanation) is first introduced in "The theory of local beables." This paper is also where he generalises his approach to include stochastic theories, which is the setting on which we shall focus here.² In part, the aim of this paper was to make explicit some notions that Bell felt were already implicit in ordinary quantum mechanics. The job of his "beables" was to recast quantum mechanics in terms of physical properties with which we are already familiar, rather than the more abstract Hermitian operators associated with quantum observables. Bell suggests such beables should satisfy a number of desiderata, but most importantly beables should correspond to something "physical", in order to distinguish variables that can be associated with "real physical" values from those that pertain to abstracta. For Bell, the latter ought to be excluded *tout court* from any causal considerations. [p57]

As an example:

²See Brown and Timpson (2014) and Wiseman and Cavalcanti (2015) for a discussion on the relationship between the deterministic background assumptions of Einstein et al. (1935), Bell (1964) and this later work of Bell's.

The beables must include the setting of switches and knobs on experimental equipment, the currents in coils and the reading of instruments. [p57]

Of particular concern are *local* beables: variables that can be assigned to a particular space-time region. Local *causality* is then defined with respect to such local beables. *Prima facie*, the idea of beables aligns fairly nicely with Pearl’s causal relata: classical random variables.

Bell goes on to define local causality in terms of these beables by introducing his famous factorisability condition. Consider two space-like separated beables, A and B associated with space-like separated regions 1 and 2. Let N denote a complete specification of all the beables belonging to the overlap of the backward light cone of 1 and 2. Let λ and M refer to the beables in the remainder of the two light cones for A and B respectively. Then, if one assumes that the joint probability $(A, B|\lambda, M, N)$ *factorises* into $(A|\lambda, M, N)(B|\lambda, M, N)$, one can use expectation values to derive an inequality (a version of the CHSH inequality), which is violated by quantum mechanics.³ For Bell, this factorisation property “says simply that correlations between A and B can only arise because of common causes N ”. For many, factorisation here *just is* Bell’s notion of local causality.

The overall idea then, is that even by adding in putative hidden beables that may exist in the joint past of beables A and B , one *still* cannot explain their correlation in locally causal terms. Thus the ‘incompleteness’ escape from non-locality suggested by EPR is blocked.

In the final section of ‘The theory of local beables’, Bell considers that despite the suggestion that nature is causally non-local, we nonetheless cannot *use* such non-locality to signal faster than light. By separating beables into two classes, those apt for human manipulation (e.g. settings) and those that are not manipulable (e.g. the outcomes), he shows that in “this *human* sense relativistic quantum mechanics is locally causal” [p64]. That is, according to quantum theory we are forbidden from manipulating one variable to induce changes in a different, space-like separated variable. Of course, this “human sense” of causation is pretty close to what interventionists use to characterise causal relations. Although, as we saw in Chapter 1, it is not simply a matter of changing one variable *here*, and seeing if another variable changes *there*, but

³An example of this is considered below.

rather causal relations are thought to be relative to a number of other specific assumptions.

In the closing paragraphs of “The theory of local beables” Bell reminds us of one further assumption required to derive the inequality: marginal independence of the settings.

It has been assumed here that the settings of instruments are in some sense free variables - say at the whim of the experimenters – or in any case not determined in the overlap of the backward light cones. Indeed, without such freedom I would not know how to formulate *any* idea of local causality, even the modest human one.

For the interventionist, such “free variables” correspond to intervention variables. The key point is not that they are somehow uncorrelated with *anything*, but rather that they are not directly correlated with any variables in the model other than their causal target. Woodward’s four criteria for an ideal intervention and Pearl’s “surgical interventions” neatly capture this requirement.

It should, by now, be fairly obvious that several of Bell’s 1976 assumptions fit nicely with the interventionist framework of Chapter 1. Interestingly, his concerns regarding the the *a priori* distinction between “controllable” and “uncontrollable” variables and also his worry that causation may require a notion of “agency” are still live debates in the interventionist literature *per se* (for example, see (Price, 2013), (Woodward, 2015), (Ismael, 2015)). It is intriguing that such concerns, at least in the context of interventionist causation, seem to be pertinent to characterisations of causation, rather than due to the peculiarities of quantum mechanics.

Also relevant for our current purposes, is the final caveat Bell adds regarding his own particular characterisation of causal explanation:

Of course, the assumptions leading to [the inequality] can be challenged. Equation 22 [factorisation] may not embody *your* idea of local causality. You may feel that only the ‘human’ version of the last section is sensible and may see some way to make it more precise.

The causal modelling formalism of Chapter 5 is an attempt to pursue this route. The more precise ‘human’ version of causation is a generalisation of interventionist causation, that allows

for the use of representative devices that go beyond the more familiar classical random variables. Whether this requires one to abandon the notion of “beables” will be addressed in Chapter 5.

We next look at a specific example of the CHSH inequality below to bring the association between Bell’s characterisation of local causality and modern interventionist causal methods into sharper focus.

2.2.1 *The CHSH inequality*

In Bell’s 1990 paper “La nouvelle cuisine” he reiterates many of the points of his earlier work. He expresses his belief that causality should not be reduced to merely “no superluminal signalling”, and also his worry that factorisability is, by itself, an inadequate representation of our causal intuitions. In this final paper, Bell uses an example of the Clauser-Horne-Shimony-Holt inequality to illustrate his views. We now consider an example of this experiment from the perspective of interventionist causal theory.

Two parties, Alice and Bob, are able to perform local measurements on a physical system received from a distant source. The same source is used for the systems received by both parties, and the two systems are emitted simultaneously. The parties randomly choose from one out of two possible measurements (the “settings”).⁴ These measurements display one out of two possible *outcomes*, labelled x and y . A first pass at a plausible DAG structure might be that of Figure 2.1.

⁴In the discussion in the following section, Glymour considers the three setting case, Wood and Spekkens consider the two setting CHSH inequality. The differences are irrelevant for what I say here.

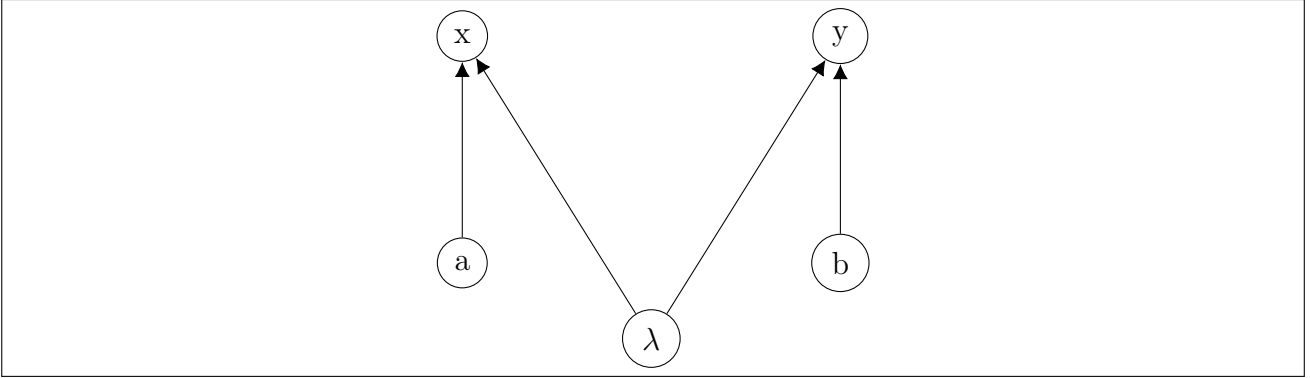


Figure 2.1: A possible DAG for the Bell experiment. a and b are random variables representing the settings at each wing, and x and y are the random variables representing the corresponding outcomes. λ captures the physical conditions occurring in the joint past of the two systems (initiated from the source).

The aim is to produce a causal DAG that can explain the correlation between x and y , where correlation implies $P(x, y) \neq P(x)P(y)$. Let us start by assuming the DAG in figure 2.1 depicts the correct causal interpretation of the experiment. According to the assumptions of Chapter 1, the following probabilistic relationships are implied by the graph:

$$Pr(x|a, \lambda, y) = Pr(x|a, \lambda) \quad (2.1)$$

$$P(y|b, \lambda, x) = P(y|b, \lambda) \quad (2.2)$$

$$Pr(x|a, b, y, \lambda) = Pr(x|a, y, \lambda) \quad (2.3)$$

$$P(y|b, a, x, \lambda) = P(y|b, x, \lambda) \quad (2.4)$$

$$P(x, y|a, b, \lambda) = P(x|a, \lambda)P(y|b, \lambda) \quad (2.5)$$

$$P(x, y, a, b, \lambda) = P(x|a, \lambda)P(y|b, \lambda)P(a)P(b)P(\lambda) \quad (2.6)$$

Factorising the joint distribution over all five variables according to the CMC leads to equation 2.6. The parentless variables (a, b , and λ) are associated with marginal distributions: $P(a), P(b)$ and $P(\lambda)$. The two child nodes (x and y) are associated with conditional distributions: $P(x|a, \lambda)$ and $P(y|b, \lambda)$. The product of all five probabilities should result in the joint distribution. Recall, it is the CMC that permits the inference from *connectedness* in the graph (causal dependence, characterised by the directed edges) to statistical *dependence* (and vice versa). Faithfulness, on the other hand, licences the inference from *unconnectedness* in the

graph (causal independence) to statistical *independence* (and vice versa). Thus, the *lack* of edge between (i) the two outcomes, (ii) between each outcome and its opposite setting, and (iii) between the setting and the hidden variable λ are all implied by Faithfulness.

Equation 2.5 is often referred to in the philosophical literature as the ‘factorisation condition’. In the interventionist framework, this condition can be seen as due to the conjunction of (i) equations 2.1 and 2.2, with (ii) equations 2.3 and 2.4. Typically, in the philosophical literature, factorisation is expressed as the conjunction of so-called ‘outcome independence’ (equations 2.3 and 2.4) and ‘parameter independence’ ($P(x|a, b, \lambda) = P(x|a, \lambda)$ and $P(y|a, b, \lambda) = P(y|a, \lambda)$). This presentation and the alternative I presented above are logically equivalent ways of representing factorisation ((Maudlin, 2011)). Assuming the graph is Faithful, condition (i) is justified by the lack of edge between x and y , and condition (ii) is justified by the lack of edge between x and b , or between y and a .

The CHSH inequality is a good example of a Bell inequality that challenges these assumptions. It is the simplest Bell inequality violated by a two qubit system. The experimental setup is identical to the one depicted in the graph above. Consider, now, that the measuring devices for both Alice and Bob are equipped with an indicator that registers either 1 or -1 each time a measurement is performed.⁵ Alice and Bob run the experiment many times, keeping careful track of the choices of measurement setting on each run, and the respective outcomes. At the end of this they are able to tabulate their results and find averages for the product of the outcomes for each of the four combinations of measurement settings (a_0, b_0) , (a_0, b_1) , (a_1, b_0) , (a_1, b_1) . We denote these averages $\langle a_i b_j \rangle$.

Based on the experimental setup, we can define a computable value, S :

$$S = \langle a_0, b_0 \rangle + \langle a_0, b_1 \rangle + \langle a_1, b_0 \rangle - \langle a_1, b_1 \rangle \quad (2.7)$$

⁵One can generalise this to the continuous case, where the outcomes fall anywhere in the continuous interval -1 to 1. To keep things simple and relatively intuitive, I will stick to the discrete case here.

Where for N repetitions:

$$\langle a_i b_j \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^{\infty} x_n(a_i) y_n(b_j) \quad (2.8)$$

where $i \in \{0, 1\}, j \in \{0, 1\}$.

Let us consider a model that will make a prediction for S , based on the interventionist assumptions of chapter 1. We assume that there exist physical properties, denoted by $\lambda \in \Lambda$. In any given repetition, we also assume the source, along with any hidden variables in the shared past, fixes the physical properties λ for the pair of systems. We allow that λ can change between repetitions, and so assign a probability density to describe this set of possibilities.

$$\langle a_i b_j \rangle = \int d\lambda p(\lambda) \sum_{xy} (xy) Pr(x, y | a_i, b_j, \lambda) \quad (2.9)$$

Recall, x and y are the outcome possibilities at each wing. This formulation is in line with Bell's later formulation of local causality (Bell (1976), Bell (1990)) as the outcomes only need to be *probabilistically* determined by the setting choices a_i and b_j . This is captured by the final conditional probability term in equation 2.9. Note also that this model is very general: λ is assumed to capture any combination of physical conditions that could result in correlations between x and y . Thus λ is representative of any hidden variable in the shared past of Alice and Bob that may account for the correlations. Equation 2.9 is implied by the Causal Markov Condition: we assume the correlations between x and y can be explained by averaging over their common causes (in this case, λ).

Assuming factorisation:

$$\langle a_i b_j \rangle_{\lambda} = \sum_{xy} (xy) Pr(x, | a_i, \lambda) Pr(y | b_j, \lambda) \quad (2.10)$$

$$= \langle a_i \rangle_{\lambda} \langle b_j \rangle_{\lambda} \quad (2.11)$$

where

$$\langle a_i \rangle_\lambda = \sum_x x \Pr(x|a_i, \lambda) \quad (2.12)$$

We can define $S(\lambda)$:

$$S(\lambda) = \sum_{i,j=\{0,1\}} \langle a_i b_j \rangle_\lambda (-1)^{i,j} \quad (2.13)$$

S is generated from $S(\lambda)$ by averaging over the hidden variables λ :

$$S = \int d\lambda p(\lambda) S(\lambda) \quad (2.14)$$

Now, we have:

$$S(\lambda) = \langle a_0 \rangle \langle b_0 \rangle + \langle a_0 \rangle \langle b_1 \rangle + \langle a_1 \rangle \langle b_0 \rangle - \langle a_1 \rangle \langle b_1 \rangle \quad (2.15)$$

$$S(\lambda) = \langle a_0 \rangle [\langle b_0 \rangle + \langle b_1 \rangle] + \langle a_1 \rangle [\langle b_0 \rangle - \langle b_1 \rangle] \quad (2.16)$$

Recall, $\langle b_0 \rangle$ and $\langle b_1 \rangle$ are either $+1$ or -1 , so either $(\langle b_0 \rangle + \langle b_1 \rangle)$ or $(\langle b_0 \rangle - \langle b_1 \rangle)$ is 0, and the other is ± 2 .

So, now we can write S as an inequality

$$S(\lambda) = \langle a_0 \rangle [\langle b_0 \rangle + \langle b_1 \rangle] + \langle a_1 \rangle [\langle b_0 \rangle - \langle b_1 \rangle] \leq 2 \quad (2.17)$$

Averaging over λ , we see S is bounded by 2 and -2.

The statistics generated by a quantum version of this experiment violate this bound and generate the result $S = 2\sqrt{2}$ (the Tsirelson bound). Accordingly, one of the assumptions underlying the model is incorrect. For the Interventionist, it is generally considered that the responsibility falls to either the Causal Markov Condition or Faithfulness.

2.3 What gives? Causal Markov vs Faithfulness

Glymour (2006) and Wood and Spekkens (2014) both apply interventionist causal modelling

methods to the Bell experiments. Both conclude that to explain the statistics produced in this experiment one is forced into giving up either the Causal Markov Condition or Faithfulness. Glymour starts with a plausible graph according to the experimental set up, and shows that the statistics do not produce a joint distribution that is Markov and Faithful to this graph. Wood and Spekkens start with the list of marginal and conditional independencies and dependencies found in the data, and show that one cannot *discover* a Markovian, Faithful DAG that can generate these statistical results.

2.3.1 Relinquishing the Causal Markov Condition

Glymour initially considers that the causal structure of the experiment may be represented via the same graph as depicted in Figure 2.1. He notes that the statistics of the experiments reveal an independence between the setting on one wing and the outcome on the other (the ‘no-signalling’ independence). Thus, according to the graph, the correlations seen between outcomes must be due to the common cause λ , and ought to be rendered independent *conditional* on this common cause. Recall, this is equivalent to Bell’s condition of local causality characterised in Bell (1976). Glymour follows Mermin in his characterisation of λ : it is a variable in the shared past of the two particles sent to each wing, that determines their properties prior to measurement.

According to the CMC one can decompose the joint probability distribution, taken over all five variables in the graph $P(a, b, x, y, \lambda)$ into a product of local conditional probability distributions as per equation 2.6. However, the statistics of the experiments do not conform to such a factorisation. Glymour then takes the reader through each of the possible ways one can attempt to alter the graph in order to account for the empirical results.

First, he considers adding an edge between λ and each of the settings (see figure 2.2). He rules out this option, sometimes known as superdeterminism, by reminding us that we can locally *intervene* and set the values of the variables a and b independently of any other variable in the model. From an interventionist perspective, this choice of setting (made by ourselves, or by a randomising machine) is parametrised by a variable outside the model, and

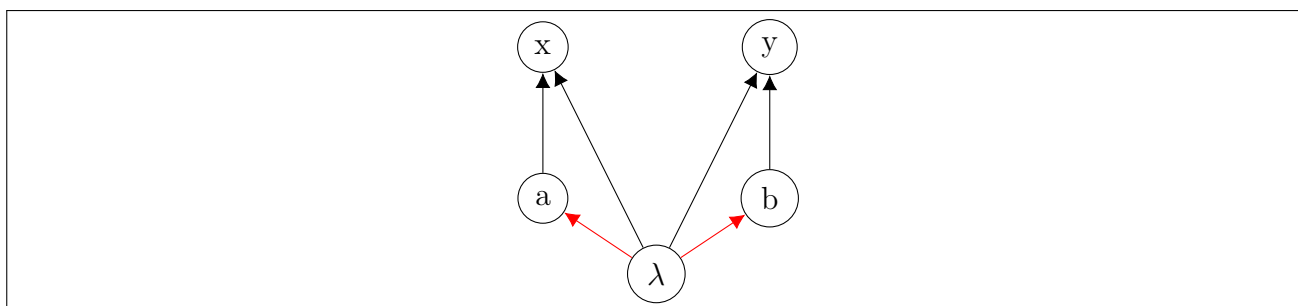


Figure 2.2: Allowing for Superdeterminism

this variable is considered to be uncorrelated with the hidden variable λ (recall, Woodward’s characterisation of an ideal intervention, section 1.3.1). In this way, any existing correlation between the measurement apparatus and λ would be broken by our intervention (recall, the effect of an intervention is to sever a variable from its incoming causal influences).

Glymour next considers the possibility of a direct causal link between the two outcomes (see figure 2.3). He notes that one can set up the experiment such that allowing for such a causal link is in direct conflict with relativity. Any causal influence between the two outcomes must propagate at superluminal speed.

The result of this analysis, according to Glymour, is that:

...real experiments...create associations that have no causal explanation consistent with the Markov assumption, and the Markov assumption must be applied to obtain that conclusion. You can say there is no causal explanation of the phenomenon, or that there is a causal explanation but it doesn’t satisfy the Markov assumption. (Glymour, 2006)[124]

His choice to consider the CMC as the target, rather than Faithfulness, is due to a proof found in Spirtes et al. (2000)[41]. For Glymour, this proof shows that the CMC implies that Faithfulness holds almost always. As we saw in the previous chapter, however, there are countless unfaithful models of natural phenomena. The point is not that such models don’t *exist*, nor that they cannot be used to represent the phenomena in question, rather the point is that typically one has the option of fine-graining and recovering a Faithful model at a new level of description. For quantum systems, it is generally considered that this option is not open to us. We come back to this point in 2.3.2.

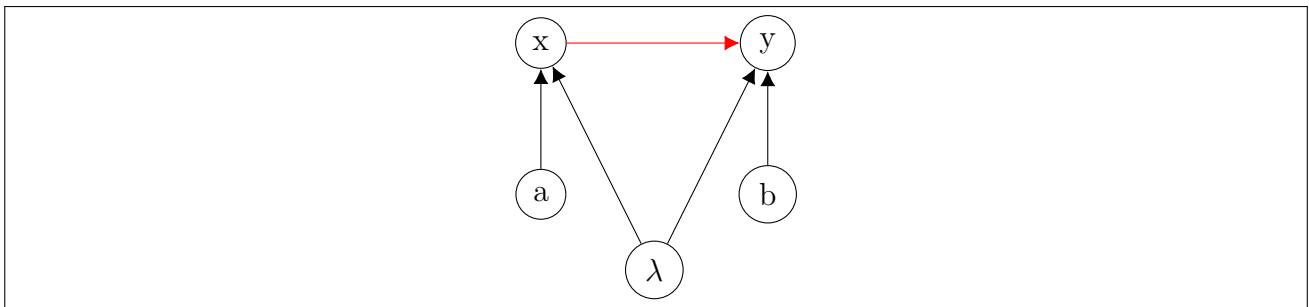


Figure 2.3: Allowing for Superluminal causes

Glymour does not offer a solution to the problems Bell experiments present for interventionist causation. Although he does offer the possibility that perhaps all causal relations do not have to satisfy the CMC, he reminds the reader that accepting this option results in another puzzle:

...why, then, does the Markov Assumption work with our experiments on middle sized dry and wet goods, with climate, and rats and drugs, and so much else?
(Glymour, 2006)[125]

We return to consider this question in Chapter 5, section 5.8.

Many other authors have also advocated relinquishing the Causal Markov Condition in order to explain the Bell correlations, although the move is not necessarily advertised as such. The ‘robustness criteria’ introduced by Redhead (1987) and debated by various philosophers in the 1980’s and 90’s, can be recast in terms of the Causal Markov Condition (Suárez and San Pedro, 2011). Roughly speaking, robustness corresponds to Pearl’s uncorrelated error variables: there may well be unmodelled causal effects, but they must remain uncorrelated for the CMC to hold. Thus, failure of robustness in the case of EPR correlations, can be recast as failure of the CMC.

Relinquishing the CMC is also often presented in the guise of modifying Reichenbach’s Common Cause Principle (RCCP): various authors have argued that weakened versions of this principle can allow for a causal explanation of quantum correlations (Hofer-Szabó et al. (2013), Leifer and Spekkens (2013), Pienaar and Brukner (2015)). Recall, the Causal Markov Condition is a generalised version of RCCP, thus to weaken RCCP to produce a causal explanation for quantum correlations implies a weakened CMC. Leifer and Spekkens (2013) suggest a modified

RCCP that implies a factorisation not of probabilities but of *conditional quantum states*. Their aim is to try and find generalised discovery algorithms that respect this altered property. Whilst this approach is promising, it remains a work in progress. We shall also see in Chapter 4 other examples of quantum causal modelling methods that advocate a weakened RCCP (for example section 4.2).

Hofer-Szabo and Vecsernyes (2012b) (hereafter HSV) suggest giving up the Law of Total Probability⁶ and define “non-commutative common causes” that also do not obey RCCP but nonetheless can be used to give a causal explanation of the Bell results. However, as Cavalcanti and Lal (2014) have shown, HSV fail to show how a common cause can act causally to provide an explanation of Bell correlations. In the HSV formalism *any* quantum product state can count as a common cause for any quantum correlation whatsoever, thus there is no causal inferential connection between common causes and their effects.

2.3.2 Relinquishing Faithfulness

Wood and Spekkens (2014) take aim at the alternative assumption. They argue that the problems of quantum causal modelling are due to the assumption of Faithfulness rather than the CMC. For these authors, the take home message is that the failure of Pearl’s methods can be explained by showing that any causal model that reproduces the observed statistics of the Bell correlations must be fine-tuned.

Using *only* marginal independence of the settings⁷ and no-signalling⁸ as input to Pearl’s discovery algorithms, Wood and Spekkens (2014) (henceforth WS) show that even if one allows for the existence of latent variables, one must permit fine-tuning of the causal parameters to both explain Bell inequality violations and still observe the no-signalling conditional independencies. They characterise this result as the conjunction of three independent theses:

1. QCORR: the observed distribution $P(a, b, x, y)$ produced by the experiment is consistent

⁶The Law of Total Probability states that for a set of mutually exclusive, exhaustive events (i.e. the individual probabilities sum to one), $\{x\}$, the probability of a distinct event y can be written in terms of the conditional $P(y|x)$, summed over the distribution for $\{x\}$: $P(y) = \sum_x P(x)P(y|x)$.

⁷ $P(a, b) = P(a)P(b)$.

⁸ $P(x|a, b) = P(x|a)$ and $P(y|a, b) = P(y|b)$.

with:

- (i) marginal independence of the settings, and
- (ii) conditional independence of the settings and opposite outcomes (no-signalling), and
- (iii) a Bell inequality is violated.

2. CAUSAL: a causal explanation of the joint distribution over the variables can be given an explanation according to Pearl's methods.
3. NOFT: the observed conditional independencies are a consequence of the Markov condition alone, and not due to fine-tuning of the causal parameters. (Wood and Spekkens, 2014)[p24]

The main result of the WS paper is that the conjunction of these three theses produces a contradiction. The proof is relatively simple: in order to explain the correlations between the observed variables (a, b, x, y) one must utilise latent variables. Adding latent variables as drains (nodes with no outgoing edges) or intermediate nodes to the graph depicted in figure 2.1 will not change the possibilities for correlations. Thus one needs only consider adding latent variables that can act as common causes. From the observed conditional independencies, $(a \perp b)$, $(x \perp b|a)$ and $(y \perp a|b)$ further independencies $((a \perp y)$ and $(b \perp x))$ can be deduced via the semi-graphoid axioms.⁹

Disallowing fine-tuning means that latent common causes can not act between a and b , or a and y , or b and x ; this would render the observed independencies $(a \perp b)$, $(a \perp y)$, $(a \perp x)$ unfaithful. So for WS, if one wishes to use causal modelling to explain Bell correlations, under the assumption of no-signalling, one must be prepared to accept fine-tuned models. No-signalling is considered not so much as an *a priori* assumption in this paper, but rather simply as a reflection of the empirical data. As we saw in Chapter 1, allowing for fine-tuned models in general will remove the possibility of causal discovery.

Recall, the standard position is that Bell showed if one wishes to *retain* causal explanation then quantum statistics forces one to accept non-locality. What Wood and Spekkens have shown is that in fact *even* if one accepts non-locality the respective causal explanation is still

⁹The semi-graphoid axioms produce the same list of independencies as using the d-separation criterion.

flawed (via the need for fine-tuning of the causal parameters). It seems we cannot save causal explanation (characterised in interventionist terms) via simply accepting non-locality.

2.3.3 *Must we choose?*

Of course, according to the paradigm of causal discovery that underpins interventionist thinking, we do not actually face a *choice* between giving up either the Causal Markov condition or Faithfulness. In Pearl's formalism, Faithfulness only carries weight as a restriction on Markovian models: if we give up Markovianity we lose the value of Faithfulness into the bargain. If we give up Faithfulness wholesale, we lose the possibility of causal discovery and the accompanying interventionist causal explanation (recall section 1.2.1). It seems neither approach affords us an interventionist causal explanation worthy of the title.

Of additional interest is the fact that none of the above authors discuss the relationship between causal structure and localised interventions. Nowhere could I find any significant discussion of the fact that the reason the CMC and Faithfulness can be considered as causal assumptions is because they produce structures that support interventionist queries. That is, structures that allow us to differentiate between effective and ineffective strategies. Keeping our eye on this broader perspective will help motivate the alternative approach I advocate in Chapter 5.

2.3.4 *Pearl and Woodward on Bell's violations*

Pearl and Woodward themselves note that their accounts cannot provide a causal explanation for quantum entanglement correlations. Woodward is content to accept that his theory suggests that there is no direct causal link between the outcomes x and y :

. . . there is no well defined notion of intervention on the spin state of one of the separated particle pairs with respect to the other in EPR type experiments. This would represent a limitation on the application of an interventionist account of causation only if there was reason to suppose that there is direct causal connection between these states. Hausman and Woodward argue there is no such reason, hence

that it is a virtue...that the interventionist account does not commit us to such a connection. (Woodward, 2007)[70]

Hausman and Woodward argue that the reason there is no well-defined notion of intervention is because one cannot regard the measurement results x and y as distinct events, or as resulting from distinct mechanisms. Quoting from Skyrms (1984) [255], they claim that the two results comprise a ‘single, indivisible non-local event’ (Hausman, 1999)[566].

There are two difficulties with this view. Firstly, it leaves the correlations between x and y causally unexplained, and secondly, as we shall see in Chapter 5, it *is* possible to intervene and break the correlation between the two measurement outcomes. Such a possibility undermines the view that the two measurement outcomes are to be thought of as a single, indivisible event.

Pearl prefers to hold fast to Laplacian determinism in order to take advantage of the intuitions that support his framework:

. . . the Laplacian conception is more in tune with human intuitions. The few esoteric quantum experiments that conflict with the predictions of the Laplacian conception evoke surprise and disbelief, and they demand scientists give up deeply entrenched intuitions about locality and causality. Our objective is to preserve, explicate and satisfy - not destroy - those intuitions. (Pearl, 2009)[26]

It is evident that Pearl considers the well-known irreducible randomness associated with quantum phenomena as a reason to consider that quantum causal mechanisms could not be deterministic. However, if causal mechanisms are characterised in the quantum case by unitary evolution, then such mechanisms will be as deterministic as those of Pearl’s more familiar classical causal mechanisms. Of course, there are key differences that emerge when one takes such evolution as representing the basic kind of causal mechanisms: we revisit this in Chapter 5.

Pearl also defends his account against various quantum qualms by claiming his causal structures apply only to macroscopic relata:

Only quantum mechanical phenomena exhibit associations that cannot be attributed to latent variables, and it would be considered a scientific miracle if anyone were

able to discover such peculiar associations in the macroscopic world (Pearl, 2009), [62]

Whilst one can quibble over what ought to count as “macroscopic”, experimentalists have pushed hard against the claim that there is some fundamental upper limit at which quantum mechanics simply ceases to apply (see, for example, Eibenberger and Tuxen (2013)). One can no longer sensibly claim that quantum mechanics applies just to the microscopic world. Nor can one sensibly hold, at least at the pragmatic level, that descriptions, explanations and models involving quantum systems are simply bereft of causal content. This perspective is defended at length in the next chapter.

2.3.5 Using interventions to discover causal relations

So why does quantum mechanics present the causal modeller with such difficulties? Perhaps the answer lies in the definition of *intervention*. It is certainly not trivial to state clearly what ought to count as an intervention in the quantum case. As we see in Chapter 4, none of the recent attempts at characterising quantum causal models have successfully addressed this worry.

Let us briefly revisit the idea of a classical intervention. One could loosely define an intervention as the fixing of a variable value, or set of values, that occurs due to a parameter outside the variable space of the model. Clearly, this definition is going to turn on what we take as a variable and value for the quantum case, ultimately a problem related to the notion of *measurement*.

An often overlooked feature of Pearl’s methods is that they abstract away from the act of measurement. An unwritten assumption of the causal modelling framework is that variable values, at least in the case of classical mechanical systems, can be considered isomorphic to physical states. Furthermore, one assumes direct access to these physical states via measurement. There are two consequences of this assumption that are worth teasing out. The first is that this grounds the distinction Pearl makes between interventionist data and observational data. Accessing observational data is assumed to be a passive endeavour: such data acquisition

(by assumption) does not interrupt the usual flow of causal information through the system being modelled. The second consequence is that the system being measured must possess non-contextual properties.

These two key assumptions presuppose the possibility of value definiteness, a luxury which quantum systems do not seem to possess. In its simplest form, value definiteness presupposes that physical systems possess properties that have definite values at all times. We assume that we can reveal these values by measuring the system at particular times and, furthermore, that these values exist independently of what we choose to measure on the system.¹⁰

So we are left with a problem. What should the interventionist take as variables and values in the case of quantum phenomena? And how can we relate these to a notion of localised intervention? Once we have answers to these questions, we can start to think more clearly about what should count as a quantum causal model. Ideally, we would have a consistent formalism that allows us to differentiate between effective and ineffective strategies involving manipulation of quantum phenomena, that also reduces to the familiar classical modelling formalism in a suitable limit. The classical version of the Causal Markov Condition and Faithfulness would then be seen as approximations of deeper quantum analogues.

This challenge is taken up in chapter 5 and represents the alternative research direction I have alluded to. By defining quantum interventions according to the possible ways in which physicists currently manipulate quantum systems, it should be possible to produce an interventionist formalism that accurately uncovers quantum causal structure. This causal structure will not necessarily conform to the intuitions of certain hidden variable theorists, who believe quantum mechanics may be an approximation of a deeper, as yet undiscovered theory. It will, however, respect current empirical evidence, and as such is still deserving of the title *quantum* causal structure.

That is *not* to say, however, that the quantum causal models produced in this thesis may not be approximations of an as yet, undiscovered, deeper, interventionist causal story. In the

¹⁰Interestingly, much has been said in the philosophy of measurement theory to dispute this claim, even for classical systems, despite its apparent simplicity (Tal, 2013). Rather than engage in this topic in any depth, however, I simply draw the readers attention to the fact that it is well known that for quantum systems, value definiteness is considered at best problematic, and at worst patently false.

light of new experimental evidence, one can simply play the game over again by producing an alternative causal theory that recovers the quantum version presented here in a suitable limit. The point is that when following this line of research one does not expect to recover “classical” interventionist models, where the mathematical objects of choice are random variables, and probabilities factorise according to the Markov condition (unless, of course, the new evidence is that we *can* signal at superluminal speeds!).

2.4 Alternative explanations

There is, of course, an alternative course of action. One can abandon the de novo search for an interventionist-style explanation of quantum correlations and instead focus on some of the existing explanations for quantum correlations we currently have on the table. I am thinking here of de Broglie-Bohm theory, Many Worlds theory and various Collapse theories. Whilst not typically characterised explicitly as *causal* explanations (perhaps with the exception of de Broglie-Bohm) there may nonetheless be the possibility for proponents of these approaches to consider how their theory may answer particular quantum interventional queries.

I have no problem with this approach (indeed I think it would be interesting to pursue this path), however, in one sense I see the metaphysical agnosticism of the quantum causal models of Chapter 5 as a virtue. If there is no known intervention that enables one to distinguish between the respective “casual models” of the more familiar interpretations, then from the perspective of interventionist causation, at least as things currently stand, they are all identical.

Furthermore, abandoning the interventionist approach to causation and causal explanation altogether comes at the cost of losing an established and successful theory of causation, a theory that *prima facie* seems surprisingly close to Bell’s own intuitions regarding causation. It seems, at least to me, that it would be preferable to have an interventionist account that can (i) preserve the key ingredients of the classical causal account, can (ii) provide a methodology for discovering quantum causal relations and (iii) can recover classical causal structure in an appropriate limit.

2.5 Summary

We have seen in this chapter that many authors have grappled with the fact that quantum casual relations cannot be shoe-horned into classical ones. Interventionist methodology fails to accurately predict the empirical results of experiments involving quantum systems. We have seen that the problem appears to lie with finding an appropriate definition of an intervention, one that can be used to determine when certain strategies will be effective and when other will not for situations where one has quantum resources. In the following chapter we take a step back and consider some motivating reasons to pursue this alternative path.

Chapter 3

Motivating Quantum Causal Models

One of the major claims of this thesis is that scientists can differentiate between effective and ineffective strategies for systems that involve quantum phenomena. Once this claim is accepted, what remains is the development of a clear formalism to help facilitate this task. In this Chapter I provide evidence for the former point.

We shall look first at how physicists represent quantum technology in order to make interventionist and counterfactual inferences: via the use of quantum circuit diagrams. Using three examples I demonstrate the manner in which these diagrams can be used to facilitate causal inferences. Two of the examples are of quantum engineered devices, and the third is a representation of a naturally occurring phenomena.¹

3.1 Introduction

As we have seen in the previous chapter, it is notoriously difficult to apply classical causal modelling techniques to systems involving quantum phenomena when we are considering certain correlations between space-like separated events. Perhaps the interventionist ought to accept that the quantum world is simply *acausal*. Such skepticism, however, presents us with a puzzle. There has been an explosion in the last decade of models, graphical representations and explanations that describe new quantum technology. These models and graphs contain counterfactual and interventionist information that facilitate a variety of seemingly causal in-

¹Part of this chapter consists of sections of my research paper (Shrapnel, 2014).

ferences. They help scientists predict, manipulate and explain. They enable physicists to know when one action will be effective and another will not. For the interventionist, what can these models be, if not causal?

The starting point for this new perspective is *not* the classical causal modelling formalism of chapters 1 and 2. Rather, the starting point is simply the formalism of quantum mechanics, defined in a suitably operational manner. Whilst it is clear that the *formalism* of interventionist causation will need to change in order to encompass these new examples, it is still possible to preserve the spirit of the account. Recall, interventionism connects with the pragmatic requirements of situated agents. In order to identify causal relationships then, what one needs to know is how to differentiate between effective and ineffective strategies. That is, to know when one course of action will lead to a particular outcome and when another will not.

3.2 Engineered quantum systems

Unlike traditional accounts of causation, the interventionist account of causation was largely inspired by casual modelling in engineering and the special sciences. As Jenann Ismael puts it, in her usual succinct manner:

Intuitions play almost no role in this [interventionist] literature. The emphasis there is on providing a framework for representing causal relations in science, i.e. a formal apparatus for rendering the deep causal structure of situations...and provides normative solutions to causal inference and judgement problems. (Ismael, 2015)[2]

For the interventionist then, the existence of quantum engineered systems should immediately alert one to the possibility of causal structure. To design such systems it *must* be possible to distinguish between effective and ineffective strategies.

Consider, for example, the representational resources physicists use in order to design quantum engineered systems. Physicists have been producing technologies that harness quantum effects for decades now. Control and manipulation of quantum systems is commonplace in the laboratory, and physicists often express the relata and relations that comprise each specific design via the use of *quantum circuit diagrams*. I think it entirely reasonable to suggest that these

diagrams may represent a form of interventionist causal structure. They are acyclic and align with temporal direction. They allow physicists to predict the effect of possible interventions and also to differentiate between effective and ineffective strategies. Of course, the strategies will not involve bringing about particular, single-case outcomes, but rather involve bringing about certain probability distributions over such outcomes. However, the long tradition of probabilistic causation suggests that, in and of itself, this feature poses no immediate threat to a causal account. On balance then, there is much to suggest that if one were looking for a notion of quantum causation, quantum circuit diagrams of engineered systems would be a good place to start.

A number of physicists have recently taken this approach towards causation and produced a generalised notion of quantum circuits.² Such “circuit models” now pervade the physics literature on quantum causation. The starting point for these accounts is *not* the classical causal modelling formalism we are used to from the philosophical literature. Rather, the starting point is simply the formalism of quantum mechanics, defined in a suitably operational manner. We will take a general look at quantum circuit models in this section, and then examine three specific examples.

For the Russellian, an obvious question arises. Surely any circuit model can be equivalently described via initial states and global Hamiltonian evolution? What forces the need for a causal interpretation? While this is true, the pragmatic value of representation via quantum circuits can not be underestimated. The *modular* form of quantum circuits, where a limited number of elements can be composed in a variety of ways, facilitates high level interventionist and counterfactual reasoning. That is, quantum circuits facilitate *causal* reasoning. The equivalent Hamiltonian representation obscures the ease with which such hypothetical manipulations are possible, and in many circumstances adds an unnecessary layer of complication.³ Of course

²See, for example, Gutoski and Watrous (2007), Chiribella et al. (2009), Chaves et al. (2015).

³A small point can be made regarding the utility of circuit diagrams, particularly in facilitating causal reasoning. IBM have recently given the general public access to a five qubit quantum computer (via a cloud based GUI - *The Quantum Experience* <http://research.ibm.com/quantum>). One can simply drag and drop various logic gates (including those that result in entanglement) and drag and drop projective and/or tomographic measurement elements, in order to configure a particular circuit. The circuit is then run on a real quantum register (sitting in millikelvin temperatures at IBM), and the results provided to the user: either as probabilities over outcomes (represented graphically) or as a ray on a Bloch sphere (one for each qubit). Indeed, several theoretical groups are now utilising this platform to remotely test experiments that have never been performed

there is nothing in this argument that says that Russellians cannot just choose to *ignore* this pragmatic advantage, and hold on to causal eliminativism.

3.2.1 Quantum Circuits

The use of quantum circuit diagrams as a representational tool developed during the 1980's and '90's. Pioneered by Deutsch (1989) these diagrams enable one to engage in high level reasoning about quantum systems using intuitive graphical objects, rather than needing to resort to complex and abstract mathematics. The utility of these methods is in part evidenced by their wide application. Circuit representations can be found in quantum information settings, in quantum biology, in quantum metrology and in condensed matter physics.⁴ They enable physicists to predict, manipulate and explain.

Broadly speaking, circuit diagrams provide an abstract representation of a complex physical process, assumed to unfold linearly in time. The entire physical process is assumed to be composed of modular subsystems, interacting with each other in a predictable manner. The modular structure allows the subsystems to be manipulated in a variety of ways, enabling a variety of counterfactual and interventionist inferences. It is the final element, the *control* component of these diagrams, that renders them useful for causal reasoning. In direct analogy to the causal models of Chapter 1, this control is linked to one or more parameters *outside* the model.

There are two distinct varieties of diagram we need to distinguish here. The first is the most ubiquitous, depicted throughout Nielsen and Chuang's famous textbook (Nielsen and Chuang, 2000). This is the version we will be referring to in this chapter. The second is the more general version utilised in the following Chapters. But first, some preliminary terminology.

previously - an example is Berta et al. (2016). One of the motivations for this project was pedagogical: given the ability to intervene locally, change individual elements and then observe the consequences, one is able to build up an intuition for the causal structure of the circuits. I say *causal* because one learns to make inferences about how changing the values of gate variables results in changes in (probabilities of) outcomes. It is thus possible to get a good feel for how quantum systems interact and influence each other without even having *heard* of Hamiltonians or Schrodinger evolution. Whilst once again, there is nothing in this argument to *force* the Russellian to drop the eliminativist style approach here, it is hard to ignore the pragmatic appeal of the interventionist causal story.

⁴That is not to say that circuit diagrams are the *only* representational device used in quantum physics, just that they are useful in a variety of explanatory contexts.

Recall, a *qubit* is simply any two state (i.e. has at least two distinguishable measurement outcomes) quantum system over which we have sufficient control. Where a system has multiple measurable outcomes the qubit can be defined by labelling any two of them.⁵ If we can accurately control this qubit then we can create and verify superposition states. Whilst classical bits can assume only one of two distinct states (1 or 0), realised by mutually exclusive physical properties, a qubit has additional states, or superpositions. For example, whilst a qubit can represent an electron spin in the ground state or excited state, the electron can also be manipulated to form a superposition state, represented by a linear combination of these two basis states. Experimental techniques have improved in recent years so that for many systems this manipulation can now proceed in a controlled and quantitative manner.

The superposition states are represented in a two-dimensional complex vector space. The state of a qubit in general is represented by a complex vector $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. The states $|\psi\rangle = |0\rangle$ and $|\psi\rangle = |1\rangle$ form an orthonormal basis for the vector space, and are known as the computational basis states. These basis states describe definite measurement outcomes for the particular physical quantity we have decided we will use to define our qubit. For example, let the $|0\rangle$ state denote an electron in its ground state and $|1\rangle$ the electron in its excited state; in this case the physical quantity that defines the computational basis is the electronic energy. The coefficients (α, β) , known as *probability amplitudes*, determine (via the Born rule) the probability with which we will return a given measurement result for a measurement of the electronic energy. That is, $|\alpha|^2$ is the probability to return the 0 outcome (the electron is in its ground state, labelled $|0\rangle$), and $|\beta|^2$ is the probability to return the 1 outcome (the electron is in an excited state, labeled $|1\rangle$). If we prepare the electron in the ground state with certainty, then $\alpha = 1$ and $\beta = 0$, so $|\psi\rangle = |0\rangle$. If we then manipulate this state into the superposition state $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, we know upon measurement that we have equal likelihood of reading a 1 or 0 outcome. In the case of the electron spin qubit such manipulation is possible through the application of timed, tuned electromagnetic pulses.

Given values for the probability amplitudes at any time and the Hamiltonian of the entire

⁵It is, of course, possible to provide circuits for systems with multiple levels, so-called qudits. To keep things clear, I will stick to the two dimensional case.

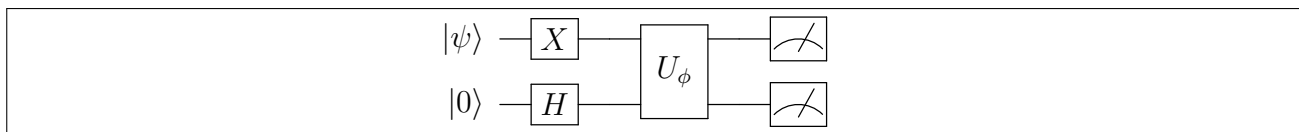


Figure 3.1: Quantum circuit. Single qubit gates X and H act on the qubits $|\psi\rangle$ and $|0\rangle$ individually. The unitary gate U_ϕ acts on both qubits and can be used to create entanglement between them. Measurements of the individual qubits are represented by the symbol on the far right of the circuit.

system, the Schrödinger equation determines the future (and past) values for the probability amplitudes. Circuit diagrams that relate qubits and gate elements inherently assume this possibility of perfect coherence.⁶

In such diagrams *wires* are associated with quantum systems whose states can be manipulated into superposition states: *logic gates* represent a physical interaction that changes the state of the quantum system in a predictable manner. As the system passes between the gates, along the wires, it is assumed that the quantum state does not change. Thus the wires in these diagrams represent the identity map, $\mathbb{1}$. The gates are assumed to effect a *unitary* evolution of the state, and can be switched on and off at various times (in figure 3.1, the boxes marked X , H and U). H and X here are just examples of particularly useful unitary gates: X swaps the state of the input (e.g. $|0\rangle \rightarrow |1\rangle$), H induces an equal superposition (e.g. $|0\rangle \rightarrow |0 + 1\rangle$ and $|1\rangle \rightarrow |0 - 1\rangle$) and U is just an example of an arbitrary unitary. The fact that such gates are unitary means that they preserve the norm of the input states. This in turn means that under the *action* of such gates quantum states will remain genuine, normalised quantum states. Measurements are typically pushed to the far right of the circuit, and represent the possible outcomes of the entire process.

These structures are acyclic and directed, and reflect possible operational choices faced by an experimenter. A specific example of such a quantum circuit is presented in Figure 3.1.

The gates in circuit diagrams are associated with matrices that transform the input states in a predictable manner. Such circuits afford a variety of counterfactual and interventionist inferences. For example, assuming one knows the input state $|\psi\rangle$ in figure 3.1, and also the

⁶This idealisation is relaxed in the more general circuits of Chapter 5.

particular value of U , it is possible to predict the distribution over the outcomes for both final measurements.⁷ Alternatively, if the unitary U is under the control of the experimenter, parametrised for example by a variable ϕ , one can make various interventionist inferences. For example, if one holds fixed the gates X and H , it is possible to manipulate the outcome probabilities in a predictable manner by altering the value of ϕ . Similarly, one can hold fixed the choice of both the unitary U and the gate X and choose to alter the gate H to an alternative, say the Z gate⁸. Again it is possible to predict the effect such a move will have on the outcome probabilities. Counterfactual reasoning follows by simple corollary: for example, one can state “*had* I changed the value of ϕ to x , the distribution over outcomes *would* have been....”.

So far, the discussion has remained rather abstract. In the next section, however, we shall see some examples of concrete implementations of such quantum circuits. It is helpful to see a few examples in order to get an intuition for the kind of causal reasoning that is possible.

3.3 Quantum Circuit Examples

There has been an interesting development in scientific explanation involving quantum mechanics over the last decade. Control and manipulation of quantum coherence in single quantum systems has become a reality in the lab environment (Dowling and Milburn (2003), Milburn (2005)). This has led to the development of many new devices capable of performing hitherto unheard of tasks. Many physicists (not least, those who develop such technology) would consider it possible to provide interventionist causal explanations for the functioning of these devices. That is, they are able to explain how these devices work and how things would be different should various parameters change. By utilising circuit diagrams, they are able to determine when certain actions will improve the functioning of the devices and when others will not.⁹

⁷In this case, knowing the input state is analogous to knowing the prior probability distribution of the exogenous (parentless) variables in a classical DAG.

⁸The Z gate adds phase with $\theta = \pi$. The $|0\rangle$ state remains unchanged and $|1\rangle$ maps to $-|1\rangle$.

⁹Russellians will no doubt look at such toy models and claim that given Hamiltonians and the Schrodinger equation they are able to evaluate exactly the same kinds of interventionist and counterfactual claims. In many circumstances however, quantum circuits are very complex. The difficulty inherent in both finding and applying Hamiltonian dynamics in these cases, makes such high level causal reasoning extremely onerous without such modular and schematic representation.

In this section we look at three examples from the scientific literature that place quantum mechanisms within complex explanatory stories. The first is an explanation for one of the novel metrological tasks now being attempted via explicit control and manipulation of a quantum system: the gravitational wave observatory, LIGO. The second is an engineered quantum system that can sense the spin state of single atoms (among other tasks): the NV diamond (Kost et al., 2014). Finally, we look at an example from ‘quantum biology’, an emerging field that identifies examples where quantum effects seem to play a functional role in biological systems. The first two examples will be used to illuminate the basics of an important technological advance, quantum metrology, and the last as an example of a naturally occurring quantum phenomenon that can be understood in interventionist terms. We will provide a rough sketch of plausible causal structures for all three, using the quantum circuit diagrams of the previous section.

At this point it is perhaps worth pausing to consider what has changed in the quantum physics community to allow quantum phenomena to now enter into such complex, *prima facie* causal, explanations. It is likely that there are many contributing factors, but possibly the most relevant breakthrough of the last few decades is physicists’ ability to isolate and probe single quantum systems. The experimental advances afforded by the work of David Wineland, Serge Haroche and colleagues have enabled the precise control and manipulation of single quantum systems, and recently earned them the 2012 Nobel Prize for Physics.

By improving our ability to isolate quantum systems, coherent quantum states in the lab can now be maintained for longer time periods, and this means that the mathematical models predicting quantum interactions and coherence times for particular set ups can be made more precise. Thus for a number of situations we can now give quantitative and relatively precise time estimates for the maintenance or loss of quantum effects such as coherence or entanglement. For such systems a close match between theoretical prediction and experimental outcome validates that the circuit model is successfully taking into account all the degrees of freedom, and the details of their interactions, that are relevant for our target explanandum (Dowling and Milburn (2003), Milburn (2005)).

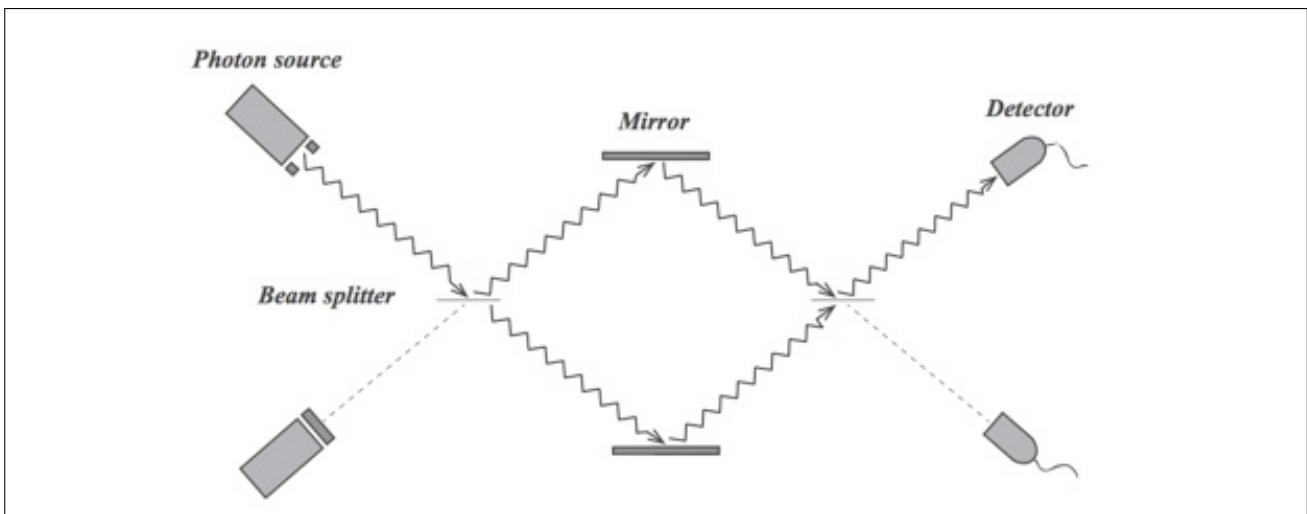


Figure 3.2: Mach-Zender Interferometer

3.3.1 Quantum optical interferometry

Optical interferometry has been used to measure length at least since the time of the famous Michelson-Morley experiment in 1887. What was not known at this time, however, is that the *sensitivity* of such devices would ultimately be determined by quantum effects. We first consider a simple quantum optical interferometer and characterise the functioning of this device via a quantum circuit. We then relate this description to the Laser Interferometer Gravitational-wave Observatory (LIGO) and its newer cousin *Advanced LIGO*.

The *classical* explanation for optical interferometry is fairly easy to understand. A laser beam is passed through a beam-splitter that divides the light into two separate beams. The two beams are then sent down the two arms of an interferometer, and reflected by mirrors at the end of each arm. Finally, the beams are recombined via a second beam splitter and measured at a photodetector. If the path lengths are exactly the same, constructive interference ensures the output light will have the same intensity as the original input laser. If the path length of one arm varies with respect to the other then the beams will be out of phase when they recombine. The consequent destructive interference results in reduced intensity of the output light. Working backwards, one can map this change in intensity to the difference in optical path length. In fig. 3.2 one can see that any event that moves the position of the uppermost mirror will alter the length of the upper arm of the interferometer. Such a disturbance can be measured by monitoring the concomitant reduction in output intensity.

LIGO is actually an enormous optical interferometer that has been designed to measure the effects of gravitational waves by monitoring the relative displacement of two mirrors in the manner described above. Gravitational waves are ripples in the fabric of space–time that are created by cosmological events such as the merging of two black holes. The measurement sensitivity of LIGO is such that we can now use it to measure displacements of around 1 part in 10^{18} : an astounding feat of modern engineering. With arms spanning around 4 km, a tiny repeated wobble is introduced into the length of each arm as gravitational waves pass perpendicularly through the device. This wobble periodically alters the relative positions of the two perfectly reflecting mirrors and thus coherently changes the intensity of the output detection events.

All optical interferometers can be given a quantum description or a classical description, and when they are operated with coherent light from a laser, the classical and quantum description of the average intensity of the light in the device are the same. Whilst this is true in LIGO as it is currently operated, the correct description of the noise properties (and thus the sensitivity) of LIGO requires a *quantum* description, even though it uses coherent states. There are two complementary sources of noise that can degrade the sensitivity of the device. The first is radiation pressure: the mirrors get a tiny momentum “kick” during the reflection of the incoming photons. As the intensity of the laser increases, so does the noise due to such radiation pressure. The second is photon counting noise (or shot noise): random variations in the departure/arrival time of the photons. These two sources of noise are not independent but linked by the quantum uncertainty principle. Thus, although one can reduce the shot noise by using high power, at sufficiently high powers, the radiation pressure noise begins to increase and limits any gain in sensitivity resulting from a decrease in shot noise. This leads to an optimum point of operation called the standard quantum limit. LIGO as it is currently run uses high power to minimise the photo detection noise and has not quite reached the standard quantum limit (SQL). In the next version of LIGO (Advanced LIGO) squeezed states will enable it to reach the SQL at lower input power. Insofar as LIGO needs to optimise its sensitivity by taking into account the Heisenberg uncertainty principle it can be considered as an example of quantum control as the

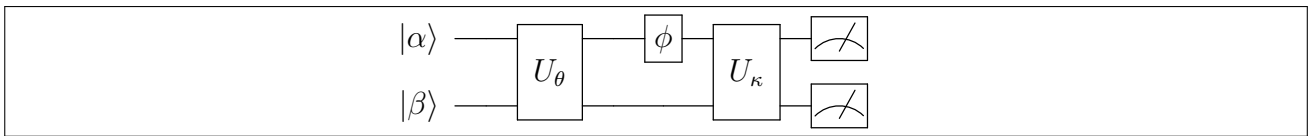


Figure 3.3: Quantum circuit for LIGO. Here U_θ represents the action of the first beamsplitter. ϕ represents the unknown phase shift introduced by the gravitational waves. U_κ represents the action of the final beamsplitter. The two measurements at the far right represent the upper and lower photodetectors. When there is no displacement of the mirror, one would expect all detection events to be in the upper port. One can use the distributions over these detection events to estimate the value of ϕ . In LIGO the input states are coherent states (photon number is poisson distributed) for the upper input, and vacuum states (no photons) for the lower input. The action of the beam-splitters for this setup will not be entangling. For Advanced LIGO, the inputs will be squeezed coherent states which *are* entangled by the beamsplitters, U_θ and U_κ

complete description requires the quantum theory of light.

Is it possible to produce a quantum circuit that represents this system? In the physics literature, there are multiple examples of optical interferometers depicted via circuit diagrams (see, for example, Caves and Shaji (2010)). For LIGO a possible quantum circuit is depicted in Figure 3.3.

We can use this very simple (and rather idealised) circuit to facilitate causal reasoning. For example, typically the values of θ and κ are fixed and associated with beamsplitters that have 50% reflectance. It is possible, however, to alter the reflectance of these mirrors and predict the accompanying change in output distributions. One can see from the diagram that altering the unitaries U_θ or U_κ will alter the distribution over outcomes for both sets of measurements.

One can also model an external intervention via control of the parameter ϕ . This is, coincidentally enough, in fact a topical issue of some import. A week after the new, updated version of LIGO opened, rumours began circulating that LIGO had detected a signal (Castelvecchi, 2015). The question was, of course, whether this represented the genuine signal of a gravitational wave. If following analysis, the signal turned out to be the kind of clean, sinusoidal wave expected to coincide with a gravitational wave detection (a ‘chirp’) then there would be two possibilities. One is that a wave had in fact been detected, the other that it was a test signal. To test the data analysis team at LIGO, the scientists will sometimes perform an elaborate drill, introducing a false signal that mirrors the kind expected to correspond to genuine wave detection. Only three members of the vast team of researchers are aware when this occurs: it

serves as an exercise to hone the analytic skills of the large research team. We now know that that the real culprit was most likely in fact two colliding black holes.

The hypothetical injection of the fake pulse can be modelled via the circuit above as a change in the value of ϕ : in the absence of any genuine waves, the experimenters have control over the value of this parameter and can intervene and introduce a false signal. Such an intervention is local, insofar as it does not change the values of the two unitaries U_θ and U_κ , and is controlled by a parameter outside the space of the model.

From this simple toy circuit it is possible to see, in principle, how one can intervene and make predictable changes to distributions over future outcomes. Counterfactual assessments are also possible: what *would* the distribution over outcomes have been *had* (for example) the reflectance of the first beamsplitter actually been different?

We now consider another quantum metrological device, and illustrate that it affords a similar representation.

3.3.2 The Nitrogen-Vacancy Diamond

Our second model describes a solid-state multi-qubit system formed by impurities in diamond, referred to as the nitrogen vacancy centre. Such systems are being used as building blocks to create more sophisticated quantum devices capable of such non-classical feats as quantum teleportation (Pfaff et al., 2014) and atomic spin sensing (Kost et al., 2014). Very recently the first ever loophole free Bell experiment was performed using this kind of device (Hensen et al., 2015).

The centre is formed by a substitutional nitrogen atom next to a missing carbon atom, with a group of electrons that cluster around the vacancy. One can represent this system as three individual qubits, associated with (i) the spin states of the electron system, (ii) the spin states of the Nitrogen nucleus, and (iii) the spin states of a nearby isotopic Carbon nucleus (see Figure 3.4).

In the ground state, the electronic system has a net spin of 1. It is a triplet system with an energy difference between the spin-0 component (the ground state) and the spin ± 1 components

(the excited states). A static magnetic field can be used to separate the spin ± 1 components, and we need only consider the -1 component. We can designate the electronic system in the ground state as $|0\rangle_e$ and in the -1 state as $|1\rangle_e$. These are the spin qubit states.

In addition, it is possible to drive the electronic system out of the ground state into an optically excited state by shining a green laser light onto the vacancy. As the electron relaxes back to the ground level it fluoresces at longer wavelengths, in the red part of the spectrum. Such fluorescence is in fact responsible for the pinkness of pink diamonds. It is this optical transition that ultimately gives us access to the spin qubit; there is more fluorescence from the $|0\rangle_e$ spin state than from the $|1\rangle_e$, and this provides a means for reading out the electronic qubit state. It is also possible to prepare (initialise) this qubit in the state $|0\rangle_e$ using a technique called optical spin pumping¹⁰.

The probability amplitudes α, β for this spin qubit can be set to specific values, i.e. the spin states of the electron can be coherently moved between the $|0\rangle_e$ state and the $|1\rangle_e$ state into specific superposition states. This is achieved by applying microwave pulses of a particular duration, carefully tuned to the frequency of the transition between the $|0\rangle_e$ and $|1\rangle_e$ state. Recall, coherence here means that the effect of the external pulse on the amplitudes is a deterministic rotation of the two dimensional complex vector (a unitary transformation).

To accurately control the spin state of the electron it is important to have precise information about the energy gap for the $|0\rangle_e$ and $|1\rangle_e$ transition. It turns out that this energy gap is extremely sensitive to local magnetic fields; its magnitude is determined both by the applied static magnetic field and by the configuration of nearby nuclei. In the NV centre the presence of a nearby nitrogen-14 and isotopic carbon-13 nucleus are the dominant nuclei that determine the finer detail of the transition frequency of the electron spin.

Because of the hyperfine interaction with these nuclear spins, the electronic excited spin state $|1\rangle_e$ is further split into sub-levels. The determination of these finer grained levels is based on the hyperfine coupling between the nuclear and electron spins; the Hamiltonian that characterises this coupling describes the magnetic dipole interaction between the electron spins

¹⁰For single shot detection, the system must operate at cryogenic temperatures to limit phonon noise (Robledo et al., 2011)

and the dominant nuclear spins. The two nuclei (carbon and nitrogen) are sufficiently far apart such that their direct magnetic dipole interaction can be neglected (approximated to zero) and we do not include this particular term explicitly in the interaction Hamiltonian, nor depict it in the circuit. The very small dipole-dipole interaction between the nuclei can simply be modelled as part of the background noise.

For most uses of the NV diamond, we can neglect the small direct coupling between the two nuclei; the interaction can be considered causally too weak to disrupt the relations of our model. However, for applications that require very long coherence times such nuclear interactions may well prove relevant. In such instances, one might be required to include the relevant term in the interaction Hamiltonian, and we then could no longer relegate this aspect of the model to background noise. For such situations, one would include a unitary gate acting between the two nuclear qubits on the circuit diagram. Note, this is directly analogous to the classical case: recall that when one changed the scope, or level of detail of the model certain latent, background influences could become causally relevant.¹¹

By applying microwaves that are scanned in frequency, and observing when we see fluorescence, it is possible to map the nuclear spins states onto the possible spin states of the electronic system, subsequently measured in a single-shot scheme. This means we also have a scheme for reading out the values of the nuclear spin states. Additionally, tuned radio-frequency pulses can individually manipulate the nitrogen and carbon spins, in a similar manner to the electron spin manipulations.

Let us review the elements of our model. We have (i) an electronic spin qubit, (ii) a nitrogen nuclear spin qubit, and (iii) a carbon nuclear spin qubit. The carbon and nitrogen nuclear spins couple to each other only very weakly. The electron spin qubit couples strongly to both nuclear spin qubits and this interaction can mediate a controllable interaction between the nuclear spins. We can intervene to alter each of the qubits via tuned, EM pulses. For the precise details of how these requirements are met for specific applications of the NV diamond see Pfaff et al. (2014), Kost et al. (2014) and Hensen et al. (2015).

A plausible simple circuit to represent the qubits and their interactions can be seen in Figure

¹¹See also section 5.10 for more on this feature of quantum causal models.

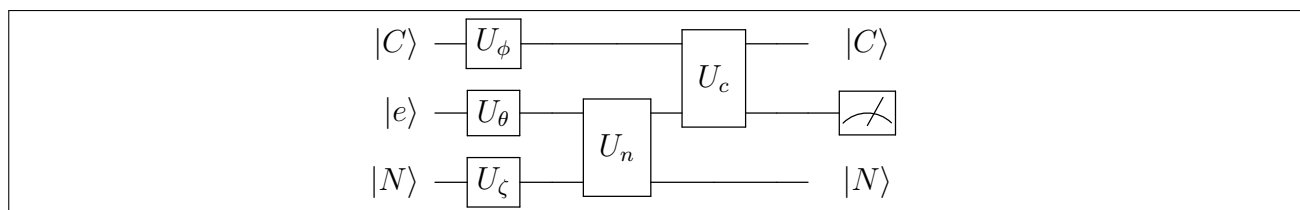


Figure 3.4: Quantum circuit for NV diamond. Here C, e and N represent the carbon, electron and nitrogen spin qubits. U_c represents the hyperfine coupling between the carbon and electron spin states, U_n represents the coupling between the nitrogen and electron spin states. The three unitaries, $U_\phi, U_\theta, U_\zeta$, represent the individual unitary control an experimenter has over each qubit, via tuned RF and microwave pulses.

4.¹²

In order to determine the spin state of a test atom, brought into close contact with the NV centre, one can imagine the presence of a third unitary acting on the electron state. By considering how the distribution over electron spin measurements changes in this modified situation, one can calculate the spin state of the test atom. One can see in the diagram above that, once again, it is possible to make various interventionist and counterfactual inferences. One can intervene from outside the space of the model and alter the various unitaries in order to test predictions relating to various hypothetical situations.

Interestingly, the kind of models and explanations we have seen in the last two examples do not just involve engineered quantum systems. They have recently been taken out of the lab, so-to-speak, and into the natural world of ‘Quantum biology’. An emerging scientific field, quantum biology also supplies models and explanations that can be given an interventionist causal interpretation. Our next example presents an illustration from this new field.

3.3.3 The European Robin

Some intriguing discoveries in biology have prompted claims that perhaps nature has already managed to do some quantum control engineering of her own (Lambert, 2012). Perhaps the two most thorough accounts are those of the avian magneto-compass (Rodgers, 2009) and the photosynthetic light-harvesting complex (Fleming, 2011). In both cases the dynamics of quantum states are thought to provide mechanisms for explaining effects that cannot be explained

¹²For the very short timescales relevant to this model, the order between the gates U_n and U_c is arbitrary.

in classical terms alone.

The current favoured explanation for the remarkable navigational abilities of the European Robin relies on the same notion of coherent quantum control that we have seen in the previous two examples. Many birds are able to navigate accurately for thousands of miles each year as they head south for the winter. Current science tells us they do so by detecting the extremely weak Earth's magnetic field lines, and the mechanism responsible is thought to rely crucially on the maintenance of quantum coherence.

The proposed mechanism is known as the radical pair mechanism, and is believed to take place in the retina of the bird. Within a molecule known as cryptochrome, the quantum spins of two electrons are initially entangled in a spin singlet state. One electron is then excited from the ground state by the absorption of an incoming photon, inducing a conformational change in the surrounding molecule and allowing the electrons to spatially separate. This separation then results in inter-conversion of the quantum state of the electron pair between a triplet and singlet state.¹³ This interconversion is considered a fully coherent, reversible quantum dynamical process, occurring due to coupling of the electron spins with separate individual nuclear environments. The dynamics of the inter-conversion are then further modulated by the earth's magnetic field.

The final step involves the electron pair encountering another reactant molecule. Depending on the joint quantum spin state of the electron pair at the time they encounter the reactant molecules, different chemical products result, and at different rates. These resultant chemicals are assumed to provide a neural signal that piggybacks onto the visual system of the bird. The speculation being that as the bird changes the orientation of its head with respect to the earth's field lines, the consequent modulation of the chemical signalling may result in a change in the distribution of light and dark patches in the bird's field of view.

The most recent behavioural experiments to support the model involve the application of weak, oscillating radio-frequency pulses to disrupt the radical pair mechanism, and consequently the birds' sense direction. These pulses are tuned to the transition frequency between triplet

¹³The singlet state = $\frac{1}{\sqrt{2}}(|10\rangle - |01\rangle)$. Initially, there are three triplet states, but hyperfine splitting due to the presence of magnetic field means that we need only consider the T_0 state = $\frac{1}{\sqrt{2}}(|10\rangle + |01\rangle)$.

and singlet states and thus disrupt the usual relationship between the joint spin state of the electron pair and the direction of the bird's head. More details can be found in (Rodgers (2009), Cai, J. (2010), Tiersch (2012), Bandyopadhyay et al. (2012), Cai, J. (2013)).

It is worth commenting that there are alternative possible explanations for the birds' remarkable navigational abilities, most notably magnetite based mechanisms. These mechanisms, however, are unable to account for the fact that (i) the compass appears to be an inclination, rather than polarity compass, (ii) the mechanism appears to be light triggered, and (iii) the birds direction sense is able to be disrupted by weak, oscillating radio-frequency pulses. For these reasons the radical pair mechanism has gained support as the favoured current explanation Lambert (2012). This means that at the very minimum, this suggests that a causal understanding is possible if the radical pair model is ultimately shown to be the correct one.

As Cai, J. (2013) recently pointed out, the explanation for the magneto- compass can be considered analogous to an optical interferometer. For the magneto-compass it is the hyper-fine coupling between the nuclear environment and electron spins that provides the process equivalent to passage through the beam splitters; formally, it is this coupling that sets up a superposition of singlet and triplet states. In the optical interferometers, interference occurs between the two paths through the interferometer, whereas in the magneto-compass it occurs in the energy eigenbasis of the electron-nuclear coupling Hamiltonian. The input and output states of the optical interferometer are the mutually exclusive photon paths; the photon is fired with certainty from either the upper or lower source, and is detected with certainty at either the upper or lower detector. In the case of the magneto-compass, the input state is a singlet state, and the detection states can be either one of the mutually exclusive states: triplet or singlet.

The final measurement step in the interferometer occurs when a photon is detected at the output; analogously, the measurement step in the avian magneto-compass occurs when the reactant molecules randomly encounter the radical pair. The changes in the probability distribution for output photons over many trials in the optical interferometer allow calculation of the minute changes of the mirror position due to passage of gravitational waves; the relative

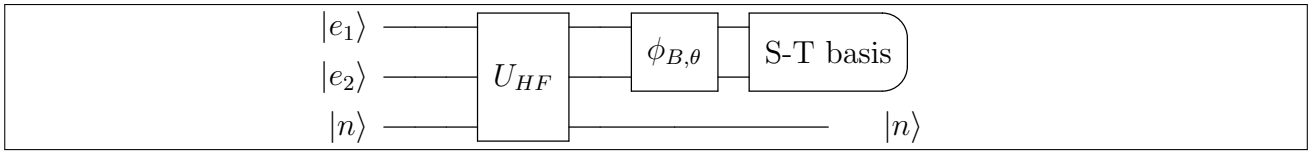


Figure 3.5: Quantum circuit for avian magnetoreception. Here $|e_1\rangle$ and $|e_2\rangle$ represent the qubits associated with the electron pair. $|n\rangle$ is associated with the spin state of the nearby nuclei. U_{HF} represents the action of the hyperfine coupling between electrons and nearby nuclei. $\phi_{B,\theta}$ represents the phase shift introduced by the Earth’s magnetic field B , parametrised by the angle with the direction of the birds head θ . Measurement is in the singlet-triplet basis.

ratios of singlet and triplet products, averaged over many molecules in the bird’s retina allows detection of the direction of the Earth’s extremely weak magnetic field. Thus the bird is able to control the relative ratio of singlet to triplet product, and thus the associated visual pattern, by altering the direction of its head with respect to the Earth’s magnetic field lines.

The dynamics of the avian model are expressed using quantum Hamiltonian dynamics, to keep track of the coherences within the system. As discussed above, detailed knowledge of coherent states can be maintained when we have sufficient knowledge and control of any degrees of freedom that influence the system of interest. In general, biological systems are complicated, wet and warm environments and were originally thought to provide too many sources of decoherence for quantum effects, such as interference, to play any functional role. However, recent improvements in the sensitivity of experimental tools means we can now detect changes at extremely short length and time scales. Recent experiments have verified that functional quantum effects (such as coherence and entanglement) can occur at very fast timescales, and are sufficiently well localised such that their effects can make a difference before slower thermal environmental effects can interfere. Furthermore, there is increasing evidence that there is often constructive interplay between thermal noise in stationary non- equilibrium systems, such that quantum correlations are not suppressed but rather enhanced or regenerated by interaction with the environment (Huelga and Plenio, 2013).

A possible circuit representation of the radical pair mechanism is depicted in figure 3.5.

Support for the avian model has come from performing various interventions, and as the interventions have become more sophisticated the model has gained strength and support from the scientific community. Scientists have considered changes in (i) the nuclei surrounding the

separated electron pair (spin state and spatial configuration): i.e changes in the value of U_{HF} (ii) the strength or direction of the external magnetic field: $\phi_{B,\theta}$, and (iii) the joint spin state of the electron pair. In the case of changes to the nuclear configurations, such interventions are modelled numerically, in the case of the magnetic field strength and directions, such interventions can be carried out in vivo. Thus, once again, the qualitative causal structure of the model is determined via intervening on various components and noting consequent changes.

3.3.4 Conclusions

In this section I have presented three examples of scientific explanations that appeal to uniquely quantum effects. I have shown that the structure of these explanations can be depicted using circuit diagrams. Such circuit models facilitate interventionist and counterfactual reasoning, allow the user to answer Woodward's so-called 'W- questions', and share many features with the causal models of Chapter 1.¹⁴ They are acyclic and align with temporal direction, are modular and possess a clear compositional structure. They also depict an element of external control that makes precise the manner in which an experimenter can intervene on an individual element of the total system, and thus interrupt its normal evolution. Quantum circuits can provide qualitative and quantitative information regarding effective and ineffective strategies for a variety of situations involving quantum phenomena. *Prima Facie*, such explanations signal the possibility of quantum interventionist causation.¹⁵

¹⁴Recall, Woodward's W-questions are of the kind 'what if things had been different?'

¹⁵The mathematical objects associated with these structures are clearly going to be very different from those of the classical causal modelling framework. In the following chapters we shall look more closely at this feature of quantum circuits.

Chapter 4

Quantum Causal Models: the state of play

It is clear from Chapter 2 that it is very difficult to characterise a quantum causal model. The relata of quantum causal models are not going to be variables in the usual classical sense. In general, one can not associate measurement outcomes with properties that exist prior to, and independently of the act of measurement or observation. The Kochen-Specker and Bell theorems both confirm the difficulties inherent in ascribing the usual local, non-contextual, classical hidden variables to represent quantum causal structure. This means one can not simply enlarge the variable set, or fine-grain the values, and recover a Markovian, Faithful causal DAG.

Having said this, we saw in Chapter 3 that there is significant motivation to consider that an interventionist account of quantum causal modelling *ought* to be possible. In this section I will examine some recent attempts to characterise quantum causal models. Due to the success of Pearl's causal modelling methods, the quantum foundations community has recently begun to examine the possibility of providing a quantum analogue. I will introduce three such accounts here. Whilst this list is not comprehensive, I believe it provides a good cross-section of the available options¹. I will follow up by comparing the three accounts and explaining why I

¹Other examples include, for instance, Tucci (1995); Chaves et al. (2015); Oreshkov and Giarmatzi (2015) and Fritz (2015).

endorse a fourth, the focus of the next chapter.

Introduction

4.1 Quantum Causal Networks

Kathryn Laskey’s 2007 paper, “Quantum Causal Networks” characterises a causal network that allows for quantum correlations. The goal is to produce an accurate representational structure for quantum correlations, rather than to uncover particular foundational or philosophical implications per se. Laskey identifies the key role interventions play in identifying *classical* causal structure and shows that local projective measurements can provide a similar function for quantum causal models. Her Quantum Causal Networks are graphical structures that can accurately represent quantum correlations.

A distinctive feature of Laskey’s approach is the use of an explicitly dualist framework. This is a consequence of her identifying entanglement correlations as *non-causal*. In order to connect nodes characterised by this kind of relation, in addition to edges that capture signalling (‘causal’) correlations, she is forced to have two different kinds of edges in her graphs.² In her formalism, graph nodes are associated with finite dimensional Hilbert spaces, and collections of such nodes can combine via the tensor product to be associated with larger dimensional Hilbert spaces. The graph structures are defined by considering possible operations on the local Hilbert spaces. This picture will become clearer as we get into the detail of the account.

Laskey assumes that quantum evolution obeys temporal ordering and consequently defines *causal* influence as occurring between time-like separated events, but not space-like separated events. As such, causal relations here are synonymous with signalling relations. Her graphs, called Sequence Associated Graphs (SAGs), capture this distinction by allowing for two kinds of edge: directed arcs for time-like separated events and undirected arcs for “contemporaneous correlations between entangled systems” [p 5]. To ensure this feature, she requires that any two pairs of nodes are related either via a directed arc, or an undirected arc, but never both. In this way, the directed arcs establish a partial order on the full set of nodes and the contemporaneous

²The more general framework I argue for in Chapter 5 avoids this kind of dichotomy.

edges partition the nodes of the graph into equivalence sets (whose elements are called CN-sets).

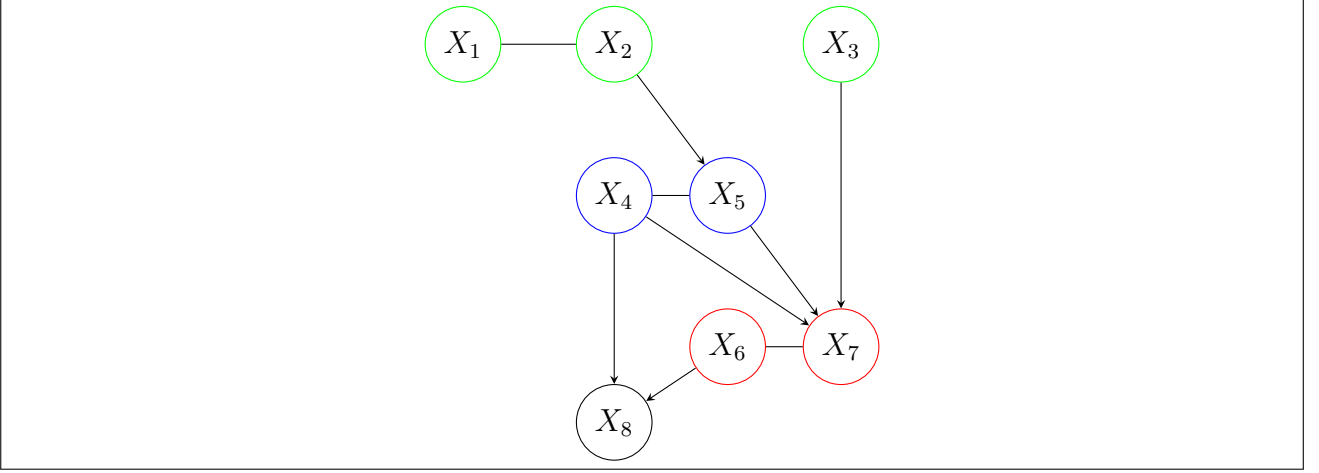


Figure 4.1: Quantum causal network. *Root* CN-sets are in green, the two *child* CN-sets are in blue and red, nodes in each CN-set are joined by undirected edges.

The CN-sets are further subdivided into two categories, in a similar manner to Pearl's endogenous and exogenous variables. *Root CN-sets* are those whose nodes have no incoming directed edges (i.e. no parents), and *child CN-sets* are those who do have such connections. Nodes that have directed edges into a member of the CN-set are known as *influencing parents* relative to that CN-set, and nodes that are contemporaneous with such influencing parents are known as *non-influencing parents*. In figure 4.1, non-influencing parents in the network are, for example, the nodes X_1 and X_7

Laskey connects the nodes of her graph with the formalism of quantum mechanics via her definition 6:

Definition 6: Let G be a SAG, let $\{X_1, \dots, X_k\}$ be a CN-set for G , and let \mathcal{H}_i denote the Hilbert space associated with X_i . Let $\{W_1, \dots, W_r\}$ denote the set of influencing and non-influencing parents for $\{X_i, \dots, X_k\}$, and let \mathcal{F}_i denote the Hilbert space associated with W_i . A *local distribution* $\Delta(\cdot)$ for $\{X_i, \dots, X_k\}$ is defined as:

1. If $\{X_i, \dots, X_k\}$ is a root CN-set, then $\Delta(X_i, \dots, X_k)$ consists of finite or a countably infinite set $\{\Delta_i(X_i, \dots, X_k)\}$ of density operators on $\mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_k$ such that $\sum_i \Delta_i(X_i, \dots, X_k)$ is equal to the identity;
2. If $\{X_i, \dots, X_k\}$ is a child CN-set, then $\Delta(X_i, \dots, X_k | W_1, \dots, W_r)$ consists of finite or a countably infinite set $\{\Delta_i(X_i, \dots, X_k | W_1, \dots, W_r)\}$ of quantum operators mapping $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_r$ to $\mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_k$, such that $\sum_i (\Delta_i(X_i, \dots, X_k | W_1, \dots, W_r))$ is trace preserving.

(Laskey, 2007)[6]

Unpacking this a little, each node (X_i) in the graph has an associated Hilbert space (\mathcal{H}_i). W is the set of all parent nodes for the particular CN-set $\{X_i, \dots, X_n\}$ under consideration. As an example, for the blue CN-set above, $W = \{X_1, X_2\}$. The *local* distribution over this set is defined by the two conditions above. Condition 1. states that if the set has *no* influencing parents, then the distribution simply consists of a set of density operators on the tensor product of the associated Hilbert spaces, \mathcal{H}_i , where these operators must sum to identity. Condition 2. states that if the CN-set *does* have influencing parents (as in this case), then the local distribution over the CN-set is given by a *conditional* distribution, consisting of a set of quantum operations. The operations here map the tensor product Hilbert space associated with *all* the parent nodes, to the tensor product Hilbert space associated with the nodes in the CN-set. The sum of the elements of this set of quantum operations must be trace-preserving.

Laskey then introduces the concept of a *fiducial reduction*. This is necessary to explicate her version of an intervention. The idea is based on Hardy's 2001 proof that shows one can characterise the behaviour of a quantum system using a finite, or countably infinite set of random variables. The proof associates *fiducial states* with *fiducial measurements*: fiducial states correspond to a set of mutually orthogonal projectors that span the state space of the system under consideration. For each fiducial state we can associate a fiducial measurement: a measurement that when applied to the appropriate fiducial state, will leave the original state unchanged. Hardy's proof showed that it is possible to characterise a quantum state accurately by specifying the outcome probabilities given each of the fiducial measurements. For Laskey, a fiducial reduction is a set of projectors in which each projector in the set is a product $F_1 \otimes \dots \otimes F_n$ of fiducial projectors (definition 7).

The SAG, G , is then defined with respect to root and child CN-sets. The local probability distribution over a *root* CN-set is said to respect G if for any fiducial reduction applied to this local distribution and any i , the conditional probability of X_i given $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$ depends *only* on the neighbours of X_i in G . This means the local probability distribution of an n -dimensional root node, can be characterised as a density matrix by specifying at most $n^2 - 1$

real numbers. The conditional probability distribution for a given *child* node, is characterised by a trace preserving quantum operation relating the output Hilbert space of the conditioning nodes to the input Hilbert space of the child node. The quantum operation for a child CN-set will require at most $n^2(m^2 - 1)$ real numbers, where m is the product of the dimensions of the nodes in the child CN-set. In general there will be various independencies in the graph, generated by the signalling structure, that reduce the number of parameters needed to specify these local distributions.

To associate the SAG with a causal structure then, Laskey defines a *quantum causal network* which is a triple $Q = (G, \{\mathcal{H}_i\}, \{\text{Pr}(\cdot)\})$, where $\{\mathcal{H}_i\}$ is a collection of Hilbert spaces, one for each node, and $\{\text{Pr}(\cdot)\}$ is a set of distributions, one for each CN-set of G . The idea is that such a graph encodes not just a model for the undisturbed evolution of a quantum system, where causal influence propagates via the directed arcs, but also a model that can accurately capture changes due to interventions.

4.1.1 Intervention

Interventions are defined in this framework as projection-valued measures, and as such have an irreducibly stochastic component. This limits interventions to “instantaneous” projectors rather than the more general class of conditional operations used in the other methods described in this chapter. Laskey states:

Although almost all treatments of interventions in the quantum theory literature are heuristic and informal, the ability to control the behaviour of quantum systems by means of interventions is an essential aspect of how quantum theory is applied in practice. The lack of a mathematically rigorous theoretical framework for analysing the effects of interventions has sowed confusion and hindered advances in practical applications of quantum theory. (Laskey, 2007)[6]

This is not quite right. In the last fifteen years there has been significant developments in formulating more rigorous notions of what ought to count as quantum control via intervention (see, for example, Peres (2000) and Wiseman and Milburn (2014)). It is possible to move

away from the use of projection-valued measured to more general kinds of measurements and operations. We shall see this in action in the next chapter (section 5.5).

4.1.2 Causal Discovery

In summary, Laskey's models provide a correct characterisation of the correlative patterns one expects for quantum systems that contain entanglement relations. There are a number of unsatisfactory elements to this presentation, however. Firstly, the use of two different edges, causal and 'merely correlational' illustrates that this version of causal modelling does not provide an understanding of causal relationships beyond signalling relations. For the interventionist, perhaps this is enough, but it leaves causal discovery a little complicated. If one is not given a *promise* that the entire network is quantum, discovery in this context would prove exceedingly cumbersome. For each CN-set one would have to search for Bell inequality violations to ensure that the entanglement correlations in question were not simply the result of an unmeasured classical common cause. And even this would not be enough: one can have entanglement without violation of a Bell inequality (Nielsen and Chuang, 2000).

This idea of accepting entanglement correlations as a brute fact, exempt from the usual requirements of causal explanation, is not new. Hausman (1999) also believed that the lesson of quantum mechanics was that we must accept the possibility of correlations that can not be explained by causal relations. He used the term *n-connected* to signify the kind of relation connecting such correlated systems: *n* is for *non-accidental* correlation. For Hausman, such connections represent a modal primitive, justified by nomological necessity. On his account (as with Laskey) correlation is due to three possibilities (i) direct cause (ii) common cause or (iii) 'mutual dependence'. The difficulty with this view, of course, is it leaves causal discovery rather unclear. How is one to *know* when one ought to stop looking for latent common causes and just accept a correlation as a brute fact?

4.1.3 Classicality

Classical causal structures would be represented in this formalism as a graph with no CN-sets. Thus the formalism affords no natural interpretation of the relation between classical and quantum causal structure, beyond the fact that for classical structure one does not need to depict the possibility of entanglement correlations.

A final possible short-coming of Laskey’s methodology is that it does not provide a natural arena for reproducing quantum circuits. Whilst it is possible to translate the quantum circuit formalism into SAG’s, such a move would be rather convoluted. The next framework we consider takes quantum circuits as a paradigm representation of quantum causal structure, and we turn to this formalism now.

4.2 A Graph Separation theorem for Quantum Causal Models

Pienaar and Brukner (2015) (henceforth PB) take as their point of departure the findings of the Wood and Spekkens (2014) paper: causal models of quantum systems require fine-tuning. They accept that classical causal modelling techniques fail for quantum systems, and set about producing an alternative framework. The aim is to find a formalism that *can* determine when two quantum events are causally related. The model is based on the quantum circuit representation, and used to define a graphical separation rule analogous to d-separation.

There is little mention of *interventions* in this framework and the overall perspective is closely aligned with Pearl: one defines causal assumptions (such as the CMC) that permit one to make causal inferences using observational data. Measurement outcomes in their framework are taken to represent observational data, and there is no characterisation of intervention.

PB take the crucial distinction between classical models and quantum models to be that intermediate variables in quantum models do not screen off. That is, whilst for classical models knowledge of a common cause renders joint effects probabilistically independent (recall figure 1.1), this is not so in the quantum case. Similarly, for a causal chain of classical variables, cause and effect variables are independent conditional on an intermediate variable, whilst this is not so in the quantum case.

In general, the approach here is very intuitive if one is familiar with the structure of quantum circuits. It is typically fairly easy to read off dependence relations from a simple circuit, as we saw in the previous chapter. PB's framework is an attempt to systematise finding probabilistic dependencies and independencies, in order to facilitate this process for larger and more complex circuits.

PB motivate their particular approach by making the fairly natural assumption that causally separated, correlated variables must share a common cause. This reflects the well known fact that independent quantum systems can only become correlated via an interaction in their common past (e.g. a common source). As such, the failure of Reichenbach's Common Cause Principle (in its probabilistic form) is here assumed to be due to a failure of factorisation: quantum common causes do not screen off. PB define Reichenbach's Common Cause principle (RCCP) as the conjunction of two distinct assumptions:

RCCP: if two variables are correlated, conditioned on the empty set, AND causally separated (neither variable is an ancestor of the other), then they are independent conditional on the set of their common causes (parents shared by both variables).

[p6]

The first assumption is that causally separated, correlated variables must share a common cause: the *principle of common cause* (PCC). The second assumption is that variables must be *screened-off* from each other by their common causes: the *factorisation principle* (FP). Thus it is the authors wish to *retain* the first assumption (PCC), which forces them to relinquish the second (FP).

Recall, classical causal models can be generated from a list of probabilistic dependencies and independencies, found in observational sample data, and associated with a graph via d-separation. In an analogous move, PB provide a list of expected probabilistic dependencies and independencies for quantum data, and provide a criterion, q-separation, that links this list to a graphical structure. Before getting to q-separation, though, we need to define the nodes and edges.

There are three distinct kinds of node in this framework, defined in terms of their incoming

and outgoing edges: (i) exogenous nodes, (ii) intermediate nodes, and (iii) drains. See figure below.

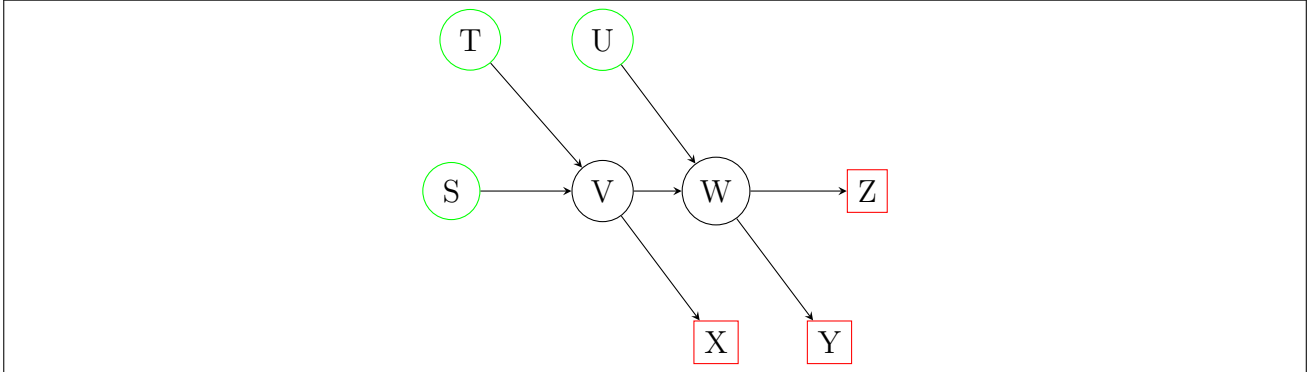


Figure 4.2: Quantum circuit DAG. *Exogenous* nodes are green and correspond to possible choices of state preparation. *Intermediate* nodes are black and correspond to possible choices of CPTP maps (in this case, unitaries). *Outcome* nodes are red and correspond to possible measurement outcomes. *Setting* nodes are circles and *outcome* nodes are squares.

(i) *Exogenous* nodes have no incoming edges, just as in the classical case. They are associated with a random variable whose value corresponds to the preparation of a normalised quantum state; formally, a density matrix. These states are restricted to exist in a Hilbert space whose dimensions equal the space formed by the tensor product of the Hilbert spaces associated with all the outgoing edges.

(ii) *Intermediate nodes*, as the name suggests, have both incoming and outgoing edges. The random variable associated with these nodes represents possible completely positive trace preserving maps (CPTP maps). Such maps can be thought of as encoding possible transformations to the state of a quantum system as it passes through the node. The fact that these maps are trace preserving, means that there are no measurement outcomes associated with the transformations taking place at these intermediate nodes. This is likely motivated by quantum circuits, where all measurements can be pushed to the far right of the circuit. We shall see some examples of this below.

(iii) *Drains* are nodes with only incoming edges. They can be used to represent either measurement *outcomes*, where the value is then associated with one out of a set of possible generalised measurements (POVMs) on the incoming systems, or *settings* where the value is associated with a CPT map, with a discarded outgoing system.

Edges are associated with a finite dimensional Hilbert space, to represent the passage of quantum systems between nodes. In circuits, these edges are wires, formally taken to represent the identity map.

Armed with these three possible kinds of node, it is possible to construct a directed acyclic graph that matches the structure of a quantum circuit. One can choose to represent the graphs by restricting the operations to pure states, unitaries and projective measurements, and consider more general operations as epistemic mixtures of these. The nice thing about the framework is it affords a very natural expression of quantum circuits in DAG form.

Recall, for quantum circuits measurements can be pushed to the far right of the circuit. Such measurement outcomes are depicted in the Quantum DAG as squares, in order to distinguish the two kinds of nodes more readily.³

The kind of probabilistic relations implied by this particular choice of nodes can be made intuitively clear by considering the physical principles we would expect for such circuits. The first observation is that *settings* ought to be distributed independently of one another, just like exogenous nodes in a classical network. In the quantum network however, these settings include *intermediate* nodes, and this fact places some significant restrictions on screening off relations. As these intermediate setting nodes will not screen off an outcome from its other ancestors, the usual Markov condition will not hold for these networks. PB show, however, that a weaker property holds: *the Quantum Causality Condition (QCC)*.

The QCC states that an outcome is independent of all *settings* that are not causes (where the cause can be either direct or indirect), and all *outcomes* that do not share a common cause, conditioned on the empty set. Sets of outcomes are not necessarily independent of each other, conditional on their common causes. They can still be dependent, even once their common causes are taken into consideration, thus violating the FP principle discussed earlier. As Butterfield (1992) noted: for quantum systems common causes do not screen off. As a result, parents do not have the same special status in this framework as they do in classical networks: causes of a variable are to be associated with *any* of its ancestors, and the usual distinction between direct and indirect causation doesn't apply.

³I use squares for outcome nodes, rather than the filled-in nodes of PB.

PB also assume an absolute ordering to the nodes (directedness) and no causal loops (acyclicity) in accordance with our usual intuitions about temporal directedness and no backwards in time causation. One can of course, relax these constraints and explore the consequences, but for this paper, the authors assume both these features.

Given these definitions of nodes, one can partition the variables into two sets: (i) outcomes and (ii) settings. Having made this distinction, it is then possible to define a *quantum input list*, \mathcal{Q}_o , as a pair $\{PA_{\mathcal{O}}, Q\}$, containing:

(i) An ordered list of parents, where each set of parents, $\mathbf{pa}(X_i)$, is a subset of the set of predecessors of X_i that are settings.

(ii) A set of conditional independence relations (Q) formed from two distinct kinds of independence:

(a) settings are independent, except when one conditions on their common outcomes.

(b) outcomes are independent of all *settings* that are not causes (where causes here equals graphical ancestors), and independent of all *outcomes* that do not share a common cause. [p16]

These constraints determine the independence relations in the graph, but a definition of the model parameters is needed to relate these structures to joint probability distributions. For a set of variables X_i , with value space \mathcal{E}_{X_i} , the quantum model parameters, F_q consist of :

(i) each *edge* is associated with a Hilbert space of finite dimension.

(ii) each outcome *drain* node X_i , is associated with a POVM, with an outcome for every value in \mathcal{E}_{X_i} .

(iii) each *exogenous* node is associated with a normalised quantum state for every value in \mathcal{E}_{X_i} .

(iv) For every intermediate node (and every setting drain node) X_i , a CPT map $\mathcal{C} : \mathcal{H}^{in} \rightarrow \mathcal{H}^{out}$ for every value in \mathcal{E}_{X_i} .

(vi) A marginal probability distribution on the value space \mathcal{E}_{X_i} of every variable that is not an outcome drain node. These marginal distributions are all mutually independent (these distributions can represent the experimenters choice of settings or the environmental conditions, or both).

Having laid the foundations, we are now in a position to define a *Quantum Causal Network* (QCN):

A QCN on a set of variables X is a pair $\{Q_{\mathcal{O}}, F_q\}$ consisting of a quantum input list for the set X , and a set of quantum model parameters F_q for the DAG G_Q generated by the input list. [p16]

So, just as a list of conditional independence relations and marginal dependencies can generate a classical DAG, the quantum input list can generate a quantum DAG G_Q . Recall, in the classical case the condition of d-separation ensures this relationship between the statistical independencies and the graph structure. PB next provide an analogous graph separation rule: *q-separation*. Following a similar form to the Verma and Pearl (1988) proof for d-separation, PB then prove q-separation to be sound and complete for quantum input lists.

4.2.1 Interventions

The authors do not discuss interventions in their framework and it is difficult to recast their formalism in terms of our usual intuitions about interventions. It is clear that one can intervene locally to set possible value of intermediate nodes, but such an action will not be arrow breaking in the usual sense. Partly this is due to the requirement that intermediate nodes represent choices of CPTP maps. Similarly, it is not clear what could count as an intervention on a measurement outcome. Having said that, from the broad perspective of interventionism, where one considers the aim is to identify the difference between actions that will bring about changes, and those that won't, the graphs do capture useful information. As far as circuits capture such information (see Chapter 3), so do the DAGs depicted here.

4.2.2 Classical Limit

PB discuss the relationship of their graphs to classical DAGs by considering if it is possible to recover a DAG by suitably modifying their QCM's to respect the Causal Markov Condition. They achieve this essentially by making all the indirect causes of outcomes direct causes. They define the classical limit for a QCM as follows:

Given a DAG G interpreted as a quantum network, the classical limit of G is a new DAG G_C obtained by

- (i) drawing a directed edge from every setting S_i to every outcome O_i that is descended from S_i in G_C , unless an edge already exists, and
- (ii) remove all edges that connect pairs of settings to each other. [p21]

Whilst it is clear that this procedure will recover a classical DAG that respects the CMC, it does not really illuminate in any interesting way the relationship between quantum and classical causal structure. One simply limits the quantum model parameters to a classical subspace and obtains a classical circuit, defined by a set of functions relating the values of outcomes given the values of their ancestors. Ancestors will become parents in the new graph G_c , and there will no longer be the possibility of intermediate nodes in the graph, thus the usual screening off properties of classical graphs will be observed.

4.3 Generalised Bayesian Networks

Henson et al. (2014)'s aim is to produce a generalised version of classical Bayesian network theory that can also accommodate non-classical resources. The non-classical resources are not restricted to just those of quantum systems. Rather, the authors wish to extend the BN framework to encompass future possible physical resources (such as, for example, so-called “PR-boxes”). In order to encompass these more generalised resources, they utilise previous work in generalised operational-probabilistic theories (OPTs)⁴. We next review some basics of this approach.

4.3.1 Generalised operational probabilistic theories

The development of OPTs was in large part inspired by the successes of quantum information theory. The development of quantum cryptography, teleportation, superdense coding and quantum metrology suggested to many physicists that there could be valuable operational consequences of taking an information-theoretic approach to quantum foundations. This led some

⁴See Chiribella et al. (2015) for a good introduction to the field of OPTs.

to consider that perhaps one could reconstruct quantum theory from some relatively intuitive, operationally motivated axioms, and in the process shed new light on various foundational issues.

More recently, Chiribella and co-workers have focussed on a specific variant: *operational probabilistic theories* (OPTs). OPTs are an extension of classical probability theory, formed by combining (i) the operational and compositional commitments of circuit models with (ii) the axioms of classical probability theory. Primitive elements are combined according to compositional features of circuits: roughly speaking, composed either ‘in series’ or ‘in parallel’. These compositional features are reflected in a graphical structure, with the mathematical validity of the approach underwritten by the category theory framework of Abramsky and Coecke (2009).⁵ An advantage of using such a general framework is that it enables the comparison of different kinds of possible physical theories. One can compare the implications of classical theory, quantum theory, stabiliser theory, gaussian quantum theory, PR boxes and more. In this manner, it is possible to tease out which features of a phenomena are specifically captured by quantum theory alone and which are due to other physical assumptions.

Operational and Compositional features of OPTs

The circuits are closely related to the circuits described in Chapter 3 (see section 3.2.1). They are defined as collections of *operations* (preparations, transformations and observations), connected by spatiotemporally propagating systems. The primitives in this theory are physically identifiable systems (A, B, C, \dots) and localisable *tests* ($\{\mathcal{C}_i\}_{i \in X}$), which represent a single use of some physical device. The systems are considered as the inputs and outputs to such tests. The elements of the tests are associated with distinguishable outcomes, labelled \mathcal{C}_i and referred to as *events*, indexed by the outcomes $i \in X$.

In cases where one wishes to denote the specific incoming and outgoing systems associated with a given test, one uses the notation \mathcal{C}_{iA}^B . If the output of one test is the same as the input system of another test, then the two tests compose in series. Alternatively, they are composed

⁵Category theory is a branch of mathematics particularly suited to structural analysis of physical systems. It is very general, and thus provides a good candidate for exploring the consequences of arbitrary physical theories. See Abramsky and Coecke (2009) for a good introduction.

in parallel. These two options are depicted graphically as:

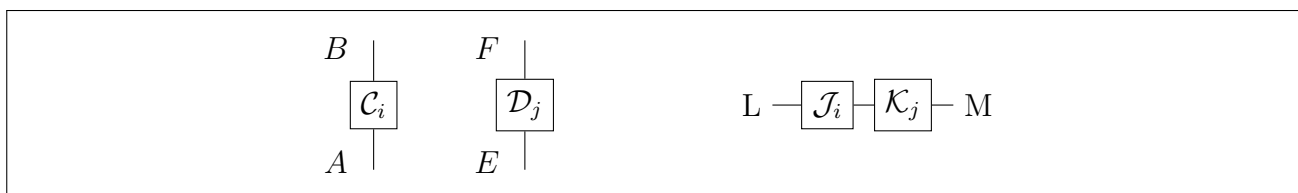


Figure 4.3: A test \mathcal{C}_i , with input system A and output system B , composed in parallel with a test \mathcal{D}_j , with input system E and output system F . Two tests, \mathcal{J}_i and \mathcal{K}_j composed in series, with input system L and output system M .

In each case, the composition yields another well-defined test, whose outcomes (i, j) are ordered pairs created by the outcomes i and j of each circuit element. For parallel compositions, the inputs and outputs also are considered together to form composite systems (AE and BF in the diagram above), and the tests also compose to form another test ($\mathcal{C}_i \otimes \mathcal{D}_j$ in the diagram).⁶

Two further primitives are required: the trivial input and the trivial output. The *trivial input system* corresponds to a preparation test, i.e it doesn't matter what the state of the input system is prior to the preparation. The *trivial output system* corresponds to a demolition measurement, i.e. it doesn't matter what the outcome state is, following the test. Tests on such systems are called *preparation-tests* and *observation-tests* respectively.

Probabilistic features of OPTs

To attach probabilities to these primitive operational elements, an OPT is defined as one in which every test from the trivial system to itself, is a probability distribution over the outcome set, and where the composition of such tests is given by the corresponding product distribution. This is depicted as a graph with no input or output wires.

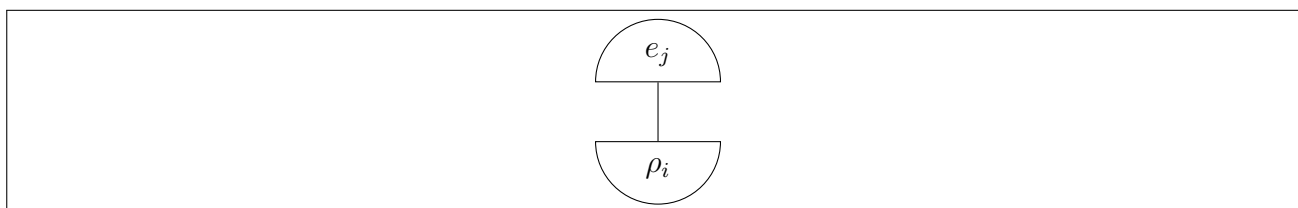


Figure 4.4: Test from trivial input to itself, associated with a preparation and then observation pair. The associated probability distribution is the joint formed by the product $p(j, i)$.

⁶The possibility of such parallel composition turns the category of events into a strict monoidal category (Chiribella et al., 2015).

An *experiment* is defined in this formalism as a sequence of tests that begins with a preparation-test and ends with an observation-test, and leaves no open wires. A simple experiment would be:

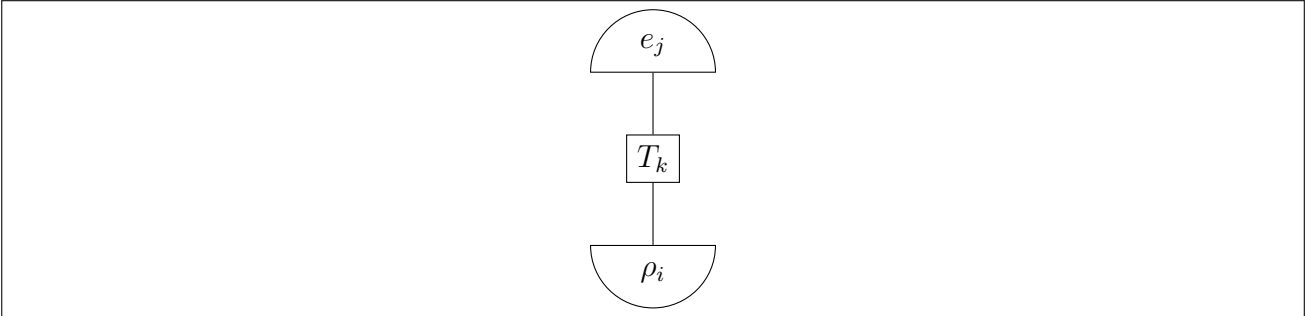


Figure 4.5: A preparation-test, transformation and then observation-test. The associated probability distribution is the joint formed by a distribution over $p(j, k, i)$ according to the probability rules of the OPT.

The Causality Criterion

The Causality Criterion is a further primitive assumption of the OPT. This criterion is stated by Chiribella et al. (2015) as the assumption that signals cannot be sent from the future to the past [p21]. One does not need explicitly to invoke this notion of temporal directedness, however, as the condition can be defined equivalently as follows. For each system A there exists a *unique deterministic effect*, T_A , where the unique deterministic effect is equivalent to the fact that summing over all possible outcomes of a test doesn't depend in any way on the input of the test. For example, in the case of quantum systems, the unique deterministic effect is characterised by the fact that POVM elements sum to the identity. The difference between preparation-tests and observation tests characterised by the causality criterion generates asymmetry in the graphs, and ultimately the directedness of the wires.

Quantum example

For quantum resources, the above elements translate in the following manner: *systems* are associated with complex Hilbert spaces and *composite* systems combine via the vector space tensor product. *Tests* are quantum instruments (completely positive maps that sum to a trace

preserving map: CPTP maps).⁷ One can think of a quantum instrument as representing a possible way in which one can interact with a quantum system, for example a choice of measurement basis, preparation, transformation etc. Preparation-tests are density matrices (unit trace positive operators) and observation-tests are $\text{Tr}(E_i)$ where $\{E_i\}$ is a POVM. The unique deterministic effect is the trace operation.

Comparing to the classical case: systems are associated with sets and composite systems are given by the cartesian product. Tests with outcome i with incoming system A and outgoing system B are given by the positive probability $P(i, \lambda_B | \lambda_A)$, where λ_A and λ_B are the possible set elements of A and B respectively. Summing over the possible outcomes i and output elements λ_B , given the input elements λ_A , equals unity. Tests compose in sequence by multiplying and summing over the λ for the intermediate system, and in parallel by multiplying. The unique deterministic effect is $P(\emptyset | \lambda)$.

The distinction between this formulation of a classical theory as an OPT and a Bayesian network lies in the fact that there are two distinct kinds of information in OPT's: systems that travel along the wires, and the classical outputs (i). This fact captures the possibility that the *outcome* of a test need not capture everything about the state of the outgoing system.

4.3.2 The Generalised Bayesian Network

The Generalised Bayesian Network is a marriage of two structures: OPTs and Bayesian Networks. Unlike traditional Bayesian networks, these networks are partitioned into two kinds of nodes, observed and unobserved. The *observed* nodes are associated with measurement outcomes (classical data), and can be associated with a classical probability distribution. In the OPT framework, observed nodes correspond to outcome-tests which output the trivial system. In the case where such observed nodes have an outgoing edge, this is taken to mean that there is a choice of test at the child node. The choice is conditioned on the classical outcome of the test at the parent node. The *unobserved* nodes are associated with 'general resources', for example the quantum source in a Bell experiment or a PR box. These unobserved nodes are *not* associated with measurement outcomes. As such, they are simply associated to a random

⁷See Davies and Lewis (1970) for an introduction to quantum instruments

variable that takes a single value with probability 1. They output *systems* but not *outcomes*. Thus it is important to remember the distinction between *output systems* and *outcomes* in this formalism. The output systems can take a variety of forms, depending on the theory under consideration (for example quantum, PR box etc) whereas the outcomes will always be classical random variables. Note, “unobserved” here doesn’t necessarily imply unknown, in the sense that one can construct a plausible graph structure, matching a particular experimental situation, by placing unobserved nodes where one would expect quantum resources.

In a similar fashion to Laskey and PB, the dual node structure of these networks demands two kinds of edges. The outgoing edges of *observed* nodes are associated with a choice of *test*, where this test is conditioned on the value of the classical value of the (observed) parent. The outgoing edges of *unobserved* nodes output only *systems*, with no associated non-trivial outcome. This is somewhat similar to the PB framework, where intermediate nodes are not associated with outcomes, although in the PB scheme, one was allowed a choice over possible intermediate node values. In the networks here, the intermediate nodes are associated with a single value that occurs with probability 1. Thus, the probability distribution associated with these generalised DAGs is generated by the distribution over the observed nodes only. Therefore, the details of the network connections between unobserved and observed nodes generate constraints on these distributions, and it is this feature that is the primary focus of the paper. For the purposes of this thesis, we will just look at the structure of these generalised Bayesian networks and ask how they fare in the context of causal discovery and in the classical limit.

The definition of the Generalised Bayesian network is the conjunction of three sub-definitions:

Definition 1. Let G be a DAG with nodes $V = \{X^{(1)}, X^{(2)}, \dots, X^{(m)}\}$. G is a generalised DAG (GDAG) if V can be partitioned into two sets of nodes:

- (i) the *observed nodes* $\{X^{(1)}, X^{(2)}, \dots, X^{(n)}\}$ (drawn as triangles), and
- (ii) the *unobserved nodes* $\{X^{(n+1)}, \dots, X^{(m)}\}$ (drawn as circles) [12]

Definition 2. Let G be a generalised DAG. Call an edge of G *observed* if it begins on an observed node, and *unobserved* if it begins on an unobserved node.

Definition 3. Let G be a GDAG with m nodes, of which the first n are observed. A probability distribution P over the observed nodes is *generalised Markov* with respect to G if there exists:

- (i) a causal operational-probabilistic theory;
- (ii) for every unobserved edge, a distinct system in the theory; and
- (iii) for every node, $X^{(i)}$, and every value of its observed parents ($Pa_o x^{(i)}$), a test denoted $\mathcal{T}_{x^{(i)}}(Pa_o x^{(i)})_{incUX^{(i)}}^{outUX^{(i)}}$.

This test is from the composite system formed by the *incoming* unobserved edges ($incUX^{(i)}$), to the composite system formed by the systems on the *outgoing* unobserved edges ($outUX^{(i)}$).

The test has:

- (a) an outcome set matching $X^{(i)}$ in the case of an observed node, but
- (b) a 1-element outcome set in the case of an unobserved node such that

$$P(x^{(1)}, x^{(2)}, \dots, x^{(n)}) = \prod_{i=1}^m \mathcal{T}_{x^{(i)}}(Pa_o x^{(i)})_{incUX^{(i)}}^{outUX^{(i)}} \quad (4.1)$$

Note that for equation (1) the joint distribution is over the observed nodes only, and generated by the product of conditionals relating inputs to outputs. The way in which these conditionals themselves are generated, depends to an extent on the theory under consideration. This is because for each particular physical theory, there will be a different mathematical operation representing the compositional element: for the quantum case it is the tensor product operation. It is, once again, easier to understand the intuition behind these definitions by examining a specific example: the Bell scenario.

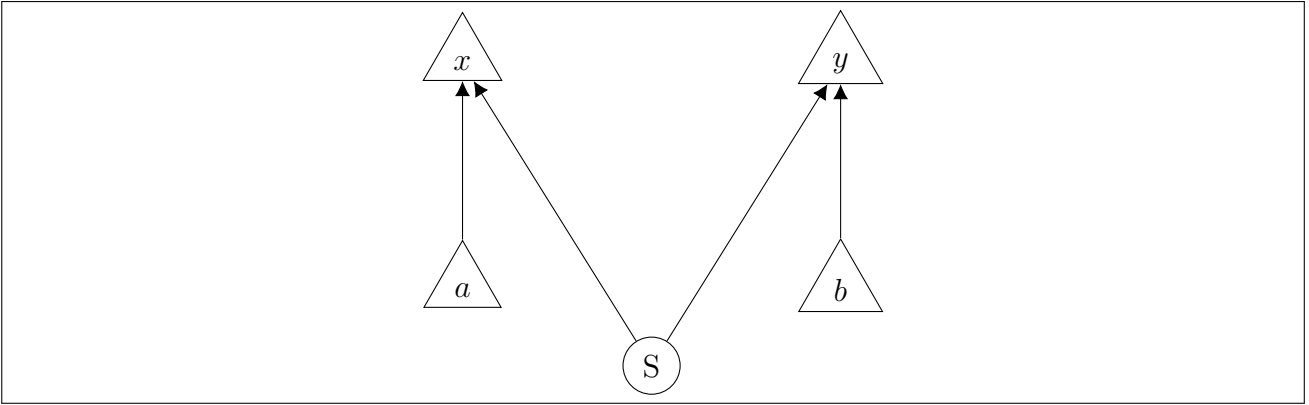


Figure 4.6: GDAG for Bell experiment. a and b are the settings at each wing and x and y are the corresponding outcomes. S is the source. a, b, x, y are all observed nodes, associated with classical probability distributions, S is an unobserved node.

In the example above, the probability that is Markov for this GDAG is

$$P(a, b, x, y) = \text{Tr}[(E_a^x \otimes E_b^y)\rho] p_a p_b \quad (4.2)$$

where E_a^x and E_b^y are POVMs of each a, b and ρ is a bipartite state.

4.3.3 D-separation

There is a definition of d-separation that holds for these graphs. As one has potential common causes in these graphs that are not classical, but rather taken from arbitrary physical resources, such unobserved nodes will again not screen off. Hence d-separation can only be defined relative to the observed nodes. Despite the apparent weakness of this criterion for independence, the final sections of the paper reveal some very interesting results. One of the primary consequences of this framework is that the causal structure defined via the GDAG formalism implies stronger constraints on the correlations than just those implied by conditional independence (in the quantum context, non-signalling) relations. Such constraints can be viewed as a generalisation of the kind implied by Bell inequalities. This proof takes us too far from our original concern: the possibility of an interventionist quantum causal model, so we will not examine this further here.

4.3.4 *Causal Discovery*

The possibility of discovery of causal structure from correlations alone is complicated by the explicit inclusion of latent variables that do not screen off. In the case where one knows the location of such latent resources, it seems likely that one can *confirm* causal structure from empirical data. However, there is no mention of interventions in this framework, and no natural method for causal discovery from statistical data alone.

4.3.5 *Classical limit*

One can recover classical causal DAGs fairly easily in this formalism by insisting that all nodes in the network are observed. Knowing when one has unobserved nodes, (quantum, classical or beyond) is of course not possible *a priori*. As such, the networks do not give us any natural intuition for the relationship between classical and quantum causal structure.

4.4 **Summary**

In this chapter I have presented three formal accounts of quantum causal models. All three provide useful insights into the possibility of quantum causal structure. None explicitly characterise what should count as a quantum intervention, or provide a formalism that allows for the discovery of causal structure. Nor do they illuminate a possible connection between classical and quantum causal structure. In the next chapter I will present a new formalism that attempts to address these two problems.

Chapter 5

Markovian Quantum Causal Modelling

We have seen that classical interventionist techniques fail to identify causal relations for quantum systems. We have also identified several motivating reasons to consider that an interventionist account of quantum causation ought to be possible. After examining some recent attempts at formalising quantum causal models, we saw they were deficient in one way or another. In this chapter I present an interventionist framework that can account for quantum causal structure (Costa and Shrapnel (2016), Shrapnel (2015)). We consider some desiderata that we would wish given the methods of classical causal modelling, and ask whether the quantum causal modelling framework presented here indeed fulfils these criteria.¹

5.1 Introduction

We saw in the last chapter that starting with quantum correlations and using classical methods to identify causal structure is problematic. It is very difficult to produce a formalism that allows for the discovery of quantum causal structure and also recovers classical causal structure in a suitable limit. In the framework presented here, we take a different approach. Starting with a plausible notion of *quantum intervention*, we consider what should count as a variable and value for the purposes of discovering quantum causal structure. It is important to remember that these structures are defined according to their ability to identify effective and ineffective

¹Part of this chapter consists of sections from my submitted paper (Shrapnel, 2015), and also parts of the collaborative research paper (Costa and Shrapnel, 2016). Any other material included here from these papers is referenced accordingly. I thank my co-author, Fabio Costa, for the opportunity to reproduce some of this work.

strategies, as presented in Chapter 3 (section 3.3.4).

At this point it is worth reminding ourselves of the desiderata we might wish the quantum causal modelling formalism should satisfy:

- (i) The formalism should allow for the discovery of causal structure from empirical data. At a minimum, such discovery should be possible using interventionist data (data instances where local events are under external control). It would be an advantage if causal structure could also be discovered in situations where interventionist information was incomplete.
- (ii) *All* correlations between empirically derived data should be accounted for via notions of direct, indirect or common cause relations, i.e. there should be no “unexplained” correlations. That is, in situations where all correlations between empirically derived data can *not* be accounted for via direct, indirect or common cause relations, there should exist a method for extending the model to include possible *unobserved* nodes in order to account for the correlations.
- (iii) Classical causal models should be recovered as a limiting case of quantum ones.

We shall see in this chapter that Quantum Causal Models (QCMs) are defined by sets of possible spatio-temporally local quantum *operations* (graphically represented as nodes) and sets of quantum *channels* that represent the causal influences acting between the nodes (graphically represented as edges). Causal structure can be *discovered* using a general notion of intervention which is defined as a choice of a quantum *instrument* (a particular set of quantum operations). Using these objects one can define direct, indirect and common cause relations. One can also show that by defining a quantum Markov condition, it is possible to identify when a causal graph is *incomplete*, in the sense that there is an unmodelled (latent) common cause. Extending the model to include such a node restores the Markov property to the causal graph. Finally, it is possible to recover classical causal structures when all local operations are diagonal in a fixed basis.

5.2 Quantum variables and values

One of the difficult tasks of developing a formalism for quantum causal modelling lies in choosing an appropriate causal variable. Recall that for classical causal models, a variable is defined by a collection of possible values, where the values often correspond to physical properties or quantities. One can think of a variable therefore as capturing a *possibility space* for a collection of mutually exclusive, labelled *events*. In the classical case, the term “events” is used in a statistical sense: an event can be spatiotemporally spread out or well localised, and in fact need not come with any particular metaphysical baggage. What is important is the relationship between possible events and (hypothetical) interventions (see section 1.2.1). That is, local interventions can choose between a range of variable values and bring about the instantiation of a particular event. Recall also, that generalised interventions are also possible and are associated with a probability distribution over events.

Recall Woodward’s pragmatic approach to identification of variables:

The problem of variable choice should be approached within a means/ends framework: cognitive enquiries can have various goals or ends and one can justify or rationalise candidate criteria for variable choice by showing that they are effective means to these ends. (Woodward, 2015a)[5]

We shall follow this advice, and ask ourselves how those physicists who routinely *use* quantum systems describe quantum events. Although there are a large variety of ways to describe quantum events, we shall choose a very general method and associate possible quantum events with *local processes*. Thus the variable space of our quantum causal models is characterised by possible quantum *operations* (or, less anthropocentrically, *processes*) that can occur inside a particular space-time region. This choice reflects our definition of a quantum intervention: a local *choice* of a set of quantum processes, where process is defined with maximal generality. Examples include the processes that can occur during measurement, preparation, transformation or even coupling to an ancilla followed by measurement. This degree of generality is needed to reflect the many possible ways in which we learn about quantum systems by performing

localised interventions.²

The mathematical object physicists associate with such general quantum operations is the trace-non-increasing completely positive map (CP map). Although a purely mathematical object, typically when a physicist *uses* such an object there is an associated operational meaning (at least in the experimental context we are considering, we see later that this restriction can be relaxed). For example, if one chooses to set a dial, change a current, tweak a mirror, in order to prepare a state or effect a transformation, then, more abstractly, such an action can be represented as choosing (probabilistically) from a range of possible CP maps. This “probabilistic choosing” is analogous to the classical, “generalised” interventions of Pearl (recall Section 1.2.1) and the “imperfect” interventions of Korb and Nicholson (2010).

So, exactly what *is* a CP map? Typically, when one thinks of the evolution of a pure quantum state one uses unitaries. Whilst unitary operators are needed to map pure states linearly to pure states, CP maps map the more general density operators to density operators.³ More carefully, a map \mathcal{M} is *positive* if it maps positive operators to positive operators, and *completely positive* if adding an ancilla system and extending the map to act on the combined Hilbert space ($\mathcal{M} \otimes \mathcal{I}$) still maps positive operators to positive operators. This means a CP map can represent evolution of a quantum state, even in the presence of entanglement.⁴ We shall look at these objects in more detail below.

It is an assumption of the framework that such quantum events can always be associated to localised spatio-temporal regions: such regions will form the nodes of our model. We also assume the particular spatio-temporally localised regions associated to the nodes have space-like boundaries. Following Oeckl (2003), we identify the past and future space-like boundaries of the region with *input* and *output* Hilbert spaces respectively (see figure 5.1)⁵. This division of the total Hilbert space into input and output spaces is crucial to the structure of the model.

²Note this is the feature missing from the three quantum causal modelling accounts of the previous chapter: the ‘intermediate’ nodes of the PB framework are not associated with measurement outcomes, and the unobserved nodes of the GDAGs formalism are likewise not associated with outcomes.

³Recall, a density operator is a positive operator with trace =1. Positivity implies the operator has non-negative eigenvalues. See Timpson (2013) Appendix A for a good review of the quantum formalism that includes such objects.

⁴See Nielsen and Chuang (2000), box 8.2 [p369] for a good illustration of the requirement of complete positivity for entangled states.

⁵Figure 5.1 has been reproduced with permission from Costa and Shrapnel (2016) and was created by Costa.

The nodes of the quantum causal model are associated to sets of CP maps, defined on the tensor product of the input and output Hilbert spaces, that tell us how the state of a quantum system can be transformed by various local actions.⁶

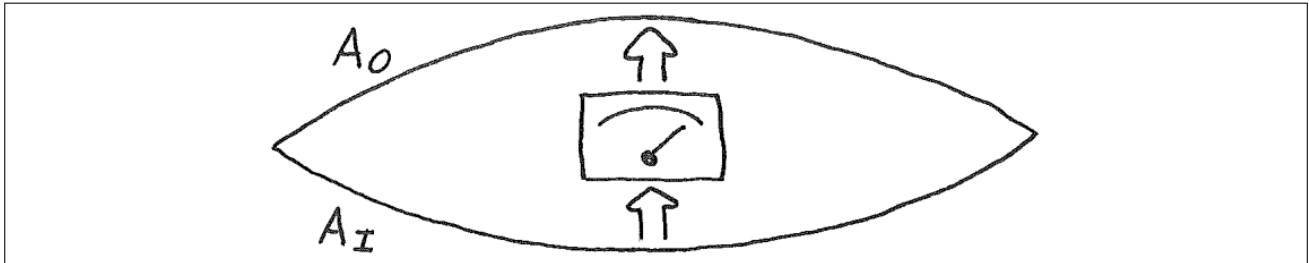


Figure 5.1: A node in a quantum causal model represents a quantum system in a region of space-time with space-like boundaries. Past and future boundaries are identified with an input and an output state space, A_I and A_O , respectively. The node is identified with the product space $A_I \otimes A_O$. A *quantum event* is a quantum operation that takes place in the local space-time region and is represented as a completely positive map from input to output space.

Recall, for classical causal models, the functional relationships between nodes are represented via directed edges. Such edges correspond to the *causal mechanisms* that are responsible for determining the statistical correlations that exist between events at different nodes. In the quantum causal models, we assume the functional relationships between the nodes are determined by quantum systems passing between different spatio-temporal regions (nodes), possibly interacting with an unknown environment. In accordance with classical models, we call such functional relationships *quantum causal mechanisms* and depict them graphically via edges. Such connecting mechanisms can *also* be represented via quantum operations, or CP maps, though they are considered to be determined by *nature*, in contrast to the maps defined at the nodes, which can be controlled. This, again, is in direct analogy with the classical case. We look at the formal details of the CP maps associated with quantum events (nodes) and the CP maps associated with causal mechanisms (edges) more carefully below.

Intuitively, one can think of a Quantum Causal Model (QCM) as a structure built from sets of CP maps that permit the representation of difference-making information. That is,

⁶ In Costa and Shrapnel (2016) the nodes are called ‘local laboratories’ in order to remain consistent with previous work on process matrices. Whilst the name implies some anthropocentric element, the term ‘local laboratory’ is taken to represent an *observer independent* spatio-temporal region where quantum events can take place. In direct analogy to the classical case (Price, 2013), one can argue whether it is possible to entirely remove the concept of observer, though I do not enter into such discussions here. For neutrality, I shall henceforth refer to the nodes as event-spaces, or simply ‘nodes’.

QCMs tell us how what can happen in one event space (i.e. which CP maps can be realised at a particular node) depends on what can happen in another. More carefully, the QCM can tell us when the *probability* of what can happen in one region depends on what can happen in another. Such difference-making information is accessed by considering the possibility of (i) interventions (local choices of sets of CP maps), and (ii) causal mechanisms that connect the nodes.

Recall, the quantum circuit diagrams of Chapter 3 are a representation of roughly this kind of structure, with gates as nodes and wires as edges. The analogy is better matched, however, to a more general kind of circuit: wires are no longer just the identity but associated with more general quantum channels, to include the possibility that the state can change between the gates. The gates no longer represent the action of a single possible unitary, but rather are place holders for a variety of possible interactions, including non-deterministic ones (to allow for the possibility of measurements). Thus the gates can also be associated with sets of completely positive maps, and it is this latter, more general, kind of circuit that will form the basis for the quantum causal modelling formalism.

How does one go about attaching probabilities to such a structure? The assignment of probabilities to quantum data is usually via the familiar Born Rule: $P(j) = \text{Tr}(O_j\rho)$, where O_j here is the relevant measurement operator and ρ the density matrix. Thus for a *single* quantum system, we know how to attach probabilities to accurately reflect (and predict) empirically derived data. What this rule does *not* tell us, however, is how to attach probabilities to multiple, interacting quantum systems. It is clear we shall need to generalise.

5.3 Process matrices and the generalised Born rule

The process matrix framework was originally developed by Oreshkov et al. (2012), building on the earlier work on ordered quantum networks (Gutoski and Watrous (2007) and Chiribella et al. (2009)). Oreshkov et al. (2012) were interested in developing a formalism that could describe the possibility of *indefinite* causal structure, where there is no assumption of either a fixed background space-time, nor a definite causal order. That is, the process matrix formalism

allows for the discovery of a *partial* ordering between the nodes of a network, by making an assumption of one-way signalling, but cannot be used to distinguish between direct, indirect or common cause structures (without added constraints). By contrast, we are interested in developing a framework that allows one to not only *fully* characterise a causal DAG, but also recognise when the model is *incomplete*, in the sense that there are unmeasured common causes.

Typically, philosophers are familiar with associating the probability of a particular measurement outcome to a particular POVM element.⁷ However, in this case we wish to encode not *just* the outcome of a measurement, but *also* the transformation to the state that occurs during the measurement process. So rather than use POVMs, we will associate probabilities with the more general CP maps introduced earlier.

First consider that we can impose some more structure on the CP maps associated with each node: a *particular* set of CP maps together can characterise a quantum *instrument*. Formally, an instrument is a set of CP maps ($\{\mathcal{M}_x\}_{x=1}^m$) that sums to a completely positive trace preserving (CPTP) map, where x labels the possible outcome events. Recall from Chapter 4, section 4.3.1, an instrument represents one out of a number of ways we can interact with the system. For example it may represent a choice of measurement setting, basis, preparation etc. Thus, for a single use of a given instrument, there will in general be a number of possible quantum events (represented as CP maps) that may occur. If such a set of CP maps is *complete* (i.e. one for every possible eventuality) then we know with certainty that at least one of them will occur. In such a circumstance the collection of CP maps will sum to a CPTP map. Such instruments define the possible interventions we can perform to test the causal structure of the model, and thus CP maps are the appropriate objects to use to characterise the event space of the node.⁸

The mathematical arena of the CP maps we associate to each node is a complex Hilbert

⁷Recall a POVM is a Positive Operator Valued Measure: a set of positive, semi-definite matrices that sum to the identity. For the usual Von Neuman projective measurements, the number of operators that compose to form the identity is equal to the dimension of the Hilbert space of the system being measured. This is not the case for POVMs. In fact, the number of available preparations and the number of available measurements on the system can be different, and also different to the Hilbert space dimension.

⁸Whilst the definition of a quantum instrument is relatively intuitive in the case of finite, discrete systems, it can become more complicated for the continuous case. See (Davies and Lewis, 1970) for a more formal definition of a quantum instrument that captures both situations. The generalisation to the continuous case does not threaten any of the arguments made here.



Figure 5.2: Generalised Circuit diagram. The nodes A and B represent distinct spatio-temporal regions. The Hilbert space associated with these regions is defined by the tensor product of the spaces associated with the incoming and outgoing systems, for example, $\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O}$ for region A . The *event* of the operation occurring at A is represented as j , and the event at B is represented as k . One can generalise to cases that include multiple incoming and outgoing edges, as we shall see in the examples below.

space. As stated, one can decompose this Hilbert space into the tensor product of two state spaces: one associated with the incoming system, \mathcal{H}_{A_I} , and one with the outgoing system, \mathcal{H}_{A_O} .⁹ The relevant CP map, denoted \mathcal{M} , is a linear map that sends states in the input Hilbert space \mathcal{H}_{A_I} to states in the output Hilbert space \mathcal{H}_{A_O} (one can imagine this map as representing a single quantum event, or process, occurring *inside* the box labelled A in Fig 5.2.). The occurrence of a particular CP map in a given space time region can be labelled according to the associated outcome $j = 1, \dots, n$. For example, for region A and outcome event j we denote the associated CP map as $\mathcal{M}_j^A : \mathcal{L}(\mathcal{H}^{A_I}) \rightarrow \mathcal{L}(\mathcal{H}^{A_O})$.

There are a variety of ways to represent CP maps corresponding to such quantum processes. Broadly speaking, they are differentiated by their pragmatic value: different communities make use of different representations to suit their individual purposes.¹⁰ Here, we will briefly consider three alternative representations: the system-environment representation, the operator-sum (Kraus) representation and the Choi-Jamiolkowski representation.¹¹ It is the last of the three that we will utilise in our quantum causal models, but it may help readers familiar with other representations to see how the three are related.

(i) The system-environment representation.

⁹For now, we consider the case of single input and output systems. We later generalise to multiple incoming and outgoing systems.

¹⁰See (Wood et al., 2015) for a good reference that presents the relationships between the different representations in pictorial form.

¹¹Thanks to Martin Ringbauer for assistance with these definitions.

The Stinespring dilation theorem (Stinespring, 1955) states that non-unitary evolution of a *local* system can always be described as unitary evolution of a *larger composite system* (the local system plus its environment) when one ignores (traces out) the environment. So for a system S in state ρ , coupled to an environment E , the evolution \mathcal{E} of S is characterised:

$$\mathcal{E}(\rho) = \text{tr}_E [U_{SE}(\rho \otimes \rho_E)U_{SE}^\dagger] \quad (5.1)$$

Surprisingly, Schumacher (1996) showed that to accurately simulate the effect of an arbitrarily large environment for a local system of d dimensions one does not (thankfully) need to include the *whole* environment but rather only an environment of d^2 .

(ii) The operator-sum representation.

An alternative representation considers a non-unitary evolution to be a unitary one, but with added noise. This *Kraus* representation is typically used in phenomenological models of noise in quantum systems. The relevant theorem states that a linear map \mathcal{E} is CPTP if and only if it may be written in the form:

$$\mathcal{E}(\rho) = \sum_{\alpha=1}^{d^2} K_\alpha \rho K_\alpha^\dagger \quad (5.2)$$

where

$$\sum_{\alpha=1}^{d^2} K_\alpha^\dagger K_\alpha = \mathbf{1} \quad (5.3)$$

Thus, the channel \mathcal{E} is decomposed as a sum of linear operators K_α (the Kraus operators) (if \mathcal{E} is unitary there is only one Kraus operator).

(iii) The Choi-Jamiolkowski representation.¹²

This final representation is perhaps the most obscure, but it is the one we will utilise to form the quantum causal models. It makes use of an isomorphism that connects

¹²This is also known as the process matrix representation, but we reserve this term for a slightly different use.

linear maps and linear operators (thus quantum channels and quantum states): the *Choi-Jamiołkowski isomorphism* (Jamiołkowski (1972), Choi (1975)).

This isomorphism utilises a procedure known as *vectorisation*: it is possible to turn an operator U into a vector $|U\rangle\rangle$ by stacking the columns of U on top of one another. Formally, to represent an operator U of d dimensions one can use an elementary basis set $|i\rangle\langle j|$, consisting of all $d \times d$ matrices with single element 1, and all other elements zero (we assume a computational basis here of $\{|i\rangle\}_{i=0}^{d-1}$). This vectorisation is represented:

$$U = \sum_{i,j=0}^{d-1} U_{ij} |i\rangle\langle j| \quad \Rightarrow \quad |U\rangle\rangle = \sum_{i,j=0}^{d-1} U_{ij} |j\rangle \otimes |i\rangle \quad (5.4)$$

and the corresponding density matrix is:

$$|U\rangle\rangle\langle\langle U| = \sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes U|i\rangle\langle j|U^\dagger \quad (5.5)$$

Thus one can construct a matrix Λ , which is the unnormalised quantum state with (trace d rather than trace 1) by vectorising a CPTP map \mathcal{E} .

$$\Lambda = \sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|) \quad (5.6)$$

The evolution of a state ρ under the channel \mathcal{E} is completely described by this matrix Λ

$$\mathcal{E}(\rho) = \text{tr}_I[(\rho^T \otimes \mathbf{1})\Lambda] \quad (5.7)$$

Confusingly, different notational choices for representing processes as CJ matrices (amounting to a transposition or partial transposition) exist in the literature (e.g. Leifer and Spekkens (2013)). In what follows, I shall use the convention defined in Oreshkov et al. (2012) and utilised in Costa and Shrapnel (2016). For a CP map $\mathcal{M} : A_I \rightarrow A_O$, where input and output spaces are the spaces of linear operators over input and output Hilbert spaces, $A_I \equiv \mathcal{L}(\mathcal{H})^{A_I}$, $A_O \equiv$

$\mathcal{L}(\mathcal{H}^{A_o})$ respectively (identified with the corresponding matrix spaces):¹³

$$M^{A_I A_o} = \sum_{i,j} |j\rangle\langle i|^{A_I} \otimes [\mathcal{M}(|i\rangle\langle j|)^{A_o}]^T, \quad (5.8)$$

$$\mathcal{M}(\rho)^{A_o} = [\text{tr}_{A_I} (\rho^{A_I} \otimes \mathbb{1}^{A_o}) \cdot M^{A_I A_o}]^T, \quad (5.9)$$

where $\{|i\rangle\}_{i=1}^{d_{A_I}}$ is an orthonormal basis in \mathcal{H}^{A_I} and T denotes transposition in that basis¹⁴. At this stage, the choice of basis here is arbitrary, but we shall later see that to enable causal discovery, it is convenient to use the Pauli basis to represent the maps and channels. According to Choi's theorem, \mathcal{M} is CP if and only if $M^{A_I A_o} \geq 0$. Henceforth CP maps will be identified with their Choi-Jamiołkowski representation unless otherwise stated.

To define valid instruments (in our context - interventions) at each node, the CJ matrices representing possible CP maps must satisfy positivity:

$$M_a^{A_I A_o} \geq 0 \quad (5.10)$$

and normalisation:

$$\text{tr}_{A_o} \left[\sum_a M_a^{A_I A_o} \right] = \mathbb{1}^{A_I} \quad (5.11)$$

A common example of a useful instrument is a measurement on an incoming system, followed by the re-preparation of a known state ρ . If the measurement corresponds to a POVM ($\{E_x\}_{x=1}^m$) where all the POVM elements sum to the identity ($\mathbb{1}^{A_I}$) then this instrument (intervention) is denoted:

$$Ins = \{E_x^{A_I} \otimes (\rho^{A_o})^T\}_{x=1}^m. \quad (5.12)$$

For a *single* node (where one is not interested in the outgoing system), one can use the

¹³Recall that complete positivity means that, for arbitrary dimensions of an ancillary system A' , the map $\mathcal{I}_{A'} \otimes \mathcal{M}$ transforms positive operators into positive operators, where $\mathcal{I}_{A'}$ is the identity map on A' . Thus, input and output spaces can have different dimensions d_{A_I} , d_{A_o} , as ancillas can be added or discarded.

¹⁴The convention here follows that of (Costa and Shrapnel, 2016): the superscript denotes the space in which a matrix is defined—hence M^X symbolises $M \in X$ (and \mathcal{J}^X for $\mathcal{J} \subset X$, if \mathcal{J} is a set of matrices), but not the matrix itself: $T^{AB} = M^A \otimes M^B$ means that T is the tensor product of two equal matrices M . Identity matrices are sometimes omitted: e.g. $M^A \otimes \mathbb{1}^B \equiv M^A$.

Born rule to attach probabilities to the event space of the node ($P(E_x^{A_I}) = \text{tr}(E_x \rho^{A_I})$). The next step is to consider how to extend this to provide a valid distribution over *multiple* nodes in a causally connected network, using CP maps. We then need to consider more general instruments, not only defined by POVM elements but also those that characterise the list of possibilities associated with re-preparation of the outgoing state. To understand how to attach probabilities in such situations, we first need to provide a definition for *causal mechanisms*. Such mechanisms are assumed to be responsible for generating the statistics across various nodes.

Causal mechanisms in the QCM formalism are defined as completely positive, trace-preserving (CPTP) maps from the output space of one node to the input space of another. Formally, we can also represent these maps using the CJ representation, although we distinguish them from the CJ representation associated to the local CP maps at each node, via the transpose operation.¹⁵ For the simple case of the causal mechanism depicted in fig 5.2, that maps outgoing states at A to incoming states at B , we represent the map \mathcal{T} :

$$T^{A_O B_I} = \sum_{j,l} |j\rangle\langle l|^{A_O} \otimes \mathcal{T}(|j\rangle\langle l|)^{B_I} \quad (5.13)$$

$$\mathcal{T}(\rho)^{B_I} = \text{tr}_{A_O} \left[\rho^{A_O} \cdot T^{A_O^T B_I} \right] \quad (5.14)$$

$$T \geq 0, \text{tr}_{B_I} T^{A_O B_I} = \mathbb{1}^{A_O} \quad (5.15)$$

where $T^{A_O^T B_I}$ denotes partial transposition on subsystem A_O .

Recall, in the classical case, autonomous causal mechanisms are considered to be responsible for the probabilistic correlations of a causal network, and it is the autonomy of the mechanisms that allows for the possibility of so-called ‘surgical’ interventions. This leads us to consider the relationship between interventions and mechanisms in the quantum case. A deterministic quantum mechanism in this formalism would correspond to a unitary map, relating the output space of one node to the input space of another. Recall in the classical case any unmodelled noise (the U_i ’s) is assumed to be uncorrelated, thus ensuring the autonomy of the mechanisms

¹⁵This representation is in accordance with the description given in equation 5.7.

and underpinning the relevance of the Causal Markov condition. In the quantum case, external noise leads to the use of a CPTP map, rather than a unitary map to represent the mechanism.¹⁶

Now we have some preliminary formal definitions for quantum events and quantum causal mechanisms we can see how probabilities are associated to a quantum causal network. First let us consider the simple case of two nodes, labelled A and B , as in the diagram above. We assume that the joint probability, $P(\mathcal{M}_j^A, \mathcal{M}_k^B)$, for a pair of particular maps to be realised must reflect a kind of non-contextuality of the CP maps: the joint probability should be independent of any variable associated with the choice of local instrument. That is, the joint probability that a particular *pair* of maps $\mathcal{M}_i^A, \mathcal{M}_j^B$ is realised should not be determined by the particular set of possible CP maps $(\mathcal{M}_1^A, \mathcal{M}_2^A, \dots, \mathcal{M}_n^A)$ associated to A 's operation.

The details of why we represent the CP maps in the particular CJ form used here relates primarily to mathematical convenience: it means we can represent the probabilities for two measurement outcomes in different space-time regions A and B as a bilinear function of the corresponding CJ operators:

$$P(\mathcal{M}_j^A, \mathcal{M}_k^B) = \text{Tr}[W^{A_I A_O B_I B_O} (M_j^{A_I A_O} \otimes M_k^{B_I B_O})] \quad (5.16)$$

where $W^{A_I A_O B_I B_O}$ is a matrix in $\mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O} \otimes \mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O})$, known as the *process matrix*. Comparing the usual Born rule to equation 5.16, one can see the process matrix W is analogous to the state ρ , and the tensor product of local CP maps (in CJ form - $(M_j^{A_I A_O} \otimes M_k^{B_I B_O})$) is analogous to the measurement operator. This particular choice of CJ representation ensures that this more generalised Born rule (5.16) reduces to the familiar one for states and POVMs (although keep in mind that for QCMs we use the more general CP maps rather than just states and POVMs).

The process matrix W thus provides a list of probabilities for all possible local outcomes, across all the nodes in the network (in this case, just two - A and B), where the *local* probabilities satisfy the rules of quantum mechanics. This imposes some constraints on possible valid process

¹⁶One can also use CPTP maps to describe irreducible noise such as that arising in dynamical collapse models. We do not commit to that particular interpretation here.

matrices:

(i) positivity

$$W \geq 0 \tag{5.17}$$

and (ii) the trace is equal to the dimension of the space associated to the outgoing systems:

$$\text{tr } W = d_o \tag{5.18}$$

However, there is a more useful way to think of the process matrix: it can also encode the possible connections between the nodes in the model. If we think of all the possible correlations between nodes as being solely determined by causal mechanisms acting between the nodes, then if the process matrix provides a full list of probabilities for local events it also captures such information.

That is, a Markovian W can be *generated* by taking the tensor product of all the CPTP maps representing connections between the nodes, written in their CJ form. Or, alternatively, it can be *discovered* by performing particular instruments at each node and observing the statistics of particular outcomes: these statistics tell us how the nodes are connected. The analogy to the classical case is relatively straightforward. In the classical case one can *generate* the causal structure by taking the product of local conditional probability distributions, or one can *discover* the causal structure by using local interventions to gather statistics that reveal the correct factorisation.

An important difference to keep in mind is that in the classical case, one can consider causal discovery for the purely observational case (where such interventionist data is not available): it is not possible to determine quantum causal structure in a similar manner. One way to see this is to imagine a quantum system, in an unknown state, entering a lab: if one is allowed to interact with it in any way (i.e. perform a measurement/intervention) then unless the state just *happens* to be in an eigenstate of the measurement performed, the observation will result in a transformation to the state and cannot be deemed “passive”. Thus in the QCM formalism, we do not make the distinction between interventional and observational data, although this

property of causal models is restored in the classical limit (where all operations are diagonal in a fixed basis: see Section 5.8).¹⁷ We can nonetheless, still ask questions about causal discovery in situations where interventionist data is incomplete (see Section 5.7).

The generalised Born Rule, Equation 5.16, says that the probability for a pair of maps, associated with events j and k at distinct locations A and B , depends on (i) the way the locations are connected (captured by the process matrix, W), and (ii) the maps themselves (captured by the CJ matrices $M_j^{A_I A_O}$ and $M_k^{B_I B_O}$). Requiring that the probabilities for CP maps are positive, and sum to 1 for any choice of instrument (CPTP map) defines a space of valid process matrices.

One of the nice features of process matrices is that they rule out causal loops, c.f. the acyclicity requirement for classical causal models. Causal mechanisms (channels) from an output space A_O to the input space of that same node A_I , will result in non-unit probabilities for some CPTP maps (Oreshkov et al., 2012).

To make these ideas a little more concrete, we look now to a simple example, involving three nodes and two wires. This example allows for the introduction of multiple outgoing system and the manner in which we represent such spaces.

5.4 Three node quantum causal models

Imagine one is given three boxes, labelled A, B and S , connected by wires, labelled (i) and (ii). Each box comes equipped with a set of levers to represent possible *instruments*, and a readout associated with possible *outcomes*. By gathering statistics that relate outcome results with respective instrument settings, one can produce a W matrix. The aim is to use this information to differentiate between the two possible situations depicted in figure 5.3 and 5.4.

¹⁷Several recent works have considered causal models that consider measurement of quantum systems to reveal “purely observational” data (Ried et al., 2015). The formalism used is in fact a special case (for bipartite systems) of the more general formalism produced here.

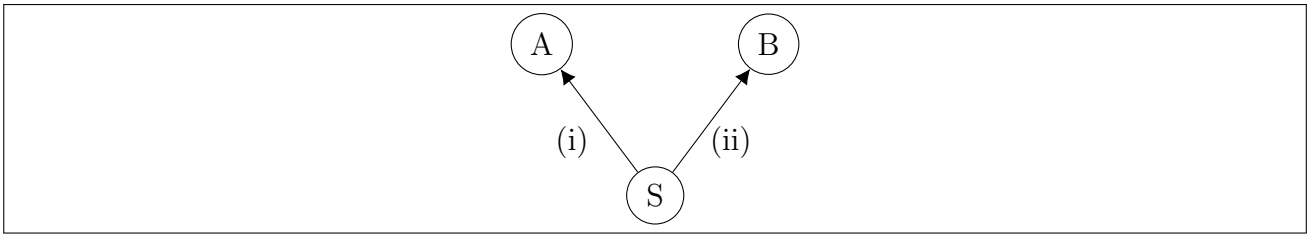


Figure 5.3: Bipartite entangled state: common cause.

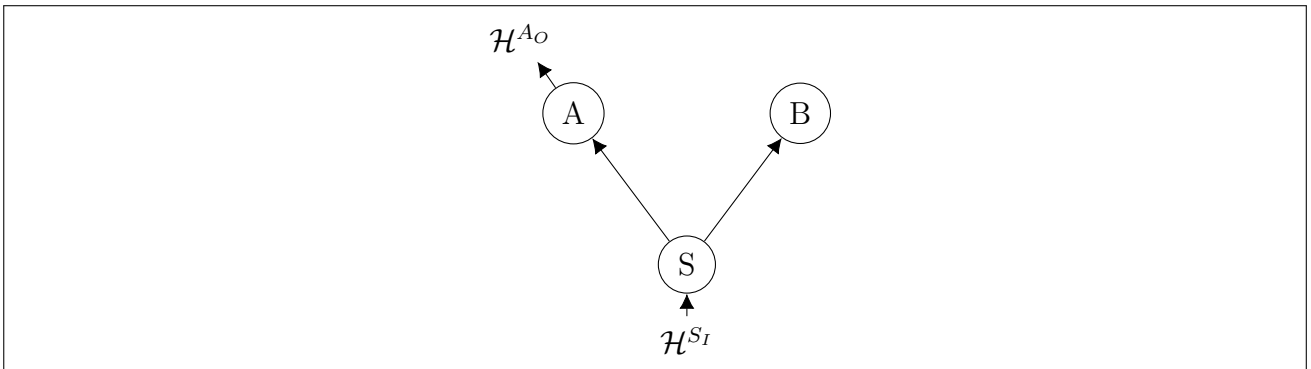


Figure 5.5: Quantum common cause structure with labelled input and output Hilbert spaces.

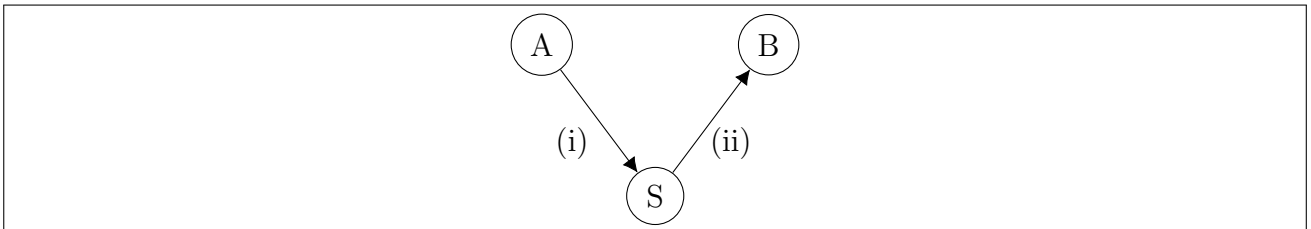


Figure 5.4: Direct cause structure.

Figure 5.3 depicts the familiar Bell scenario: the production of an entangled bipartite state. We shall refer to this as the common cause structure. Figure 5.4 depicts the direct transfer of a quantum physical system between three distinct locations: a direct causal structure. The process matrix ought to be able to tell us which of the two circuits has been implemented.

In order to depict these two structures as quantum causal models, we first need to associate an input and output space for all three nodes. In this simple case we need an input and output space for each of A and B (labelled $\mathcal{H}^{A_I}, \mathcal{H}^{A_O}, \mathcal{H}^{B_I}, \mathcal{H}^{B_O}$). For S we need one input space (\mathcal{H}^{S_I}) and two output spaces ($\mathcal{H}^{S_{O_1}}, \mathcal{H}^{S_{O_2}}$).

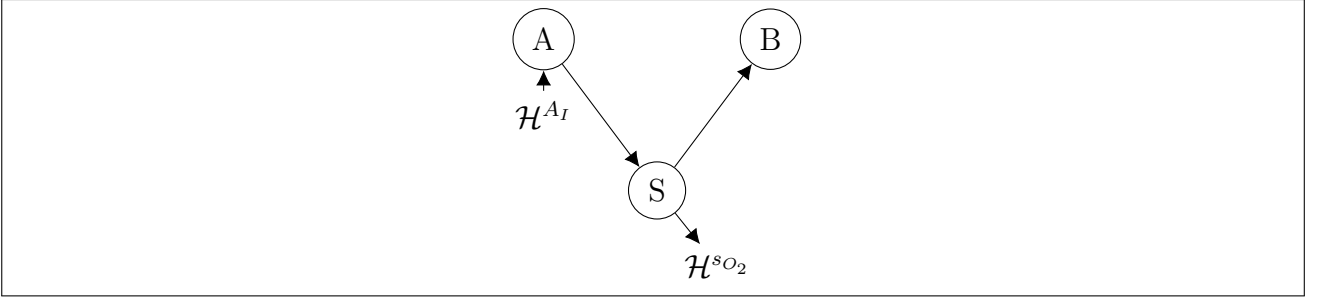


Figure 5.6: Quantum direct cause structure with labelled input and output spaces.

Each node therefore has an output space formed by the tensor product of the Hilbert spaces associated to the outgoing systems (S_e , where e labels each outgoing edge), and a *parent space* PS associated to the incoming systems. For example, the output space for node S in the common cause structure, figure 5.5, is defined by the tensor product of a tensor factor associated to each outgoing system, $O_s = S_{O_1} \otimes S_{O_2}$. The parent space for S in the direct cause structure, figure 5.6, is $PS_S = A_O$. If one were to include a third node, C providing control over the input at A , then A would have a non-trivial parent space C_O .

First let us use these example networks to compose a process matrix for each case from the terms in each example. For a *known* causal structure, one can compose the process matrix from the tensor product of terms corresponding to the various components of the network. For example, the process matrix for the common cause structure will compose as:

$$W_c = \rho^{S_I} \otimes T_1^{S_{O_1} A_I} \otimes T_2^{S_{O_2} B_I} \otimes \mathbb{1}^{A_O B_O} \quad (5.19)$$

Where ρ^{S_I} is the input state at S , $T_1^{S_{O_1} A_I}$ is the CJ matrix representing the channel from S to A , $T_2^{S_{O_2} B_I}$ is the CJ matrix representing the channel from S to B , and $\mathbb{1}^{A_O B_O}$ is the identity map on the output states at A and B . The T are analogous to the conditional distributions that describe classical causal mechanisms, and the ρ^{S_I} analogous to the prior distribution over parentless (exogenous) classical nodes. In a similar manner, one can compose the process matrix for the direct cause structure as:

$$W_d = \rho^{A_I} \otimes T_1^{A_O S_I} \otimes T_2^{S_{O_1} B_I} \otimes \mathbb{1}^{S_{O_2}} \otimes \mathbb{1}^{B_O} \quad (5.20)$$

Here ρ^{A_I} represents the input state at A , $T_1^{S_O1A_I}$ represents the connection from A to S , $T_2^{S_O1B_I}$ represents the connection between S and B , $\mathbb{1}^{S_O2}$ represents that component of the output space of S that in this case we trace out, and $\mathbb{1}^{B_O}$ is the identity on the output of B_O .

These examples show that, generally speaking, it is easy enough to decompose the process matrix into tensor products of various circuit components if we already know the structure being described. The crucial question, of course, is whether one can go in the other direction. Can one reconstruct the correct W matrix from the observed statistics of outcomes at A , B and S ? Furthermore, is it possible to differentiate between the two cases depicted and thus *discover* the causal structure of the network from that data alone?

Interestingly, if one has enough statistical data one can construct a unique circuit from a given process matrix W . As with the classical case, if one doesn't have enough data, an equivalence class of models can be recovered. So how much data is enough? This is related to the notion of “informationally complete” sets of measurements (POVMS): a single quantum state can be discovered correctly given such a set of measurements (Prugovecki, 1977). One can generalise the notion of informational completeness to extend to sets of CP maps (instruments) and situations that model multiple systems.¹⁸ As with the classical case, causal discovery in this more general setting is intimately related to the characterisation of an ideal intervention. Given enough interventions, we can fully characterise the process matrix, W and uniquely determine the causal structure of the model.

5.5 Quantum Interventions

Recall, for classical causal models, generalised interventions that select from *sets* of events are possible (section 1.2.1). Such interventions characterise distributions over possible events that can occur in a given run of an experiment. While such generalised interventions are relevant in both classical and quantum settings, the interpretation is quite different in each. In the classical case, we consider such probabilistic interventions as being due to either ignorance, or coarse-grained control: there is always the *in principle* possibility of reducing such probabilistic

¹⁸This is due to the possibility of process and state tomography (Nielsen and Chuang, 2000), and can be formally extended to the case of process matrices (Costa and Shrapnel, 2016).

interventions to ones that choose a particular event with certainty. In the quantum case, we accept that interventions of this sort will not lead to a meaningful representation of a causal model. Just as it is not possible to model *all* measurements of an arbitrary quantum state such that they yield unique results, interventions are generalised in the same way. As a result, the usual clean distinction between observational data and interventionist data can not be made: to obtain the full set of data available for quantum systems one needs to include the possibility of measurements that transform the state (recall, section 2.3.5).

It is possible to model a particular quantum instrument as an arrow-breaking intervention by performing a measurement on an incoming system, followed by the re-preparation of a known state ρ . If the measurement corresponds to a POVM $(\{E_x\}_{x=1}^m)$ where all the POVM elements sum to the identity ($\mathbb{1}^{A_I}$) then this instrument (intervention) is denoted:

$$Ins = \{E_x^{A_I} \otimes (\rho^{A_O})^T\}_{x=1}^m. \quad (5.21)$$

If one chooses equally among different state preparations, then one can remove the influence of the incoming system: an act analogous to Pearl's do-operator. In this sense, a quantum intervention can also be arrow-breaking in a similar manner to its classical counterpart and used to discover direct, indirect and common cause structures.

5.5.1 Quantum Direct Causation

For a simple two node example, it is possible to produce a definition for a direct cause. We can then generalise this to a definition that will apply to more complex situations.

Definition 4. A is a *direct cause* of B iff there exists

- (i) at least two instruments (CPTP maps) I_i^A and I_j^A associated with node A , where i and j label different instruments
- (ii) a CP map M^B associated with node B , such that

$$Tr[I_i^A \otimes M^B \cdot W^{AB}] \neq Tr[I_j^A \otimes M^B \cdot W^{AB}] \quad (5.22)$$

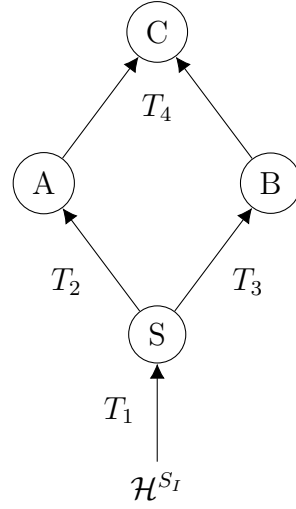
This is clearly a ‘difference-making’ definition. Practically speaking, this equation says that if there exists a possible instrument at A that can make a difference to the probability distribution over at least one of the outcomes at B , then A and B are related as cause to effect. In a similar manner to Woodward’s definition of causation, this is an existentially quantified definition. All it takes for A to be a cause of B , is for there to exist an appropriate means for intervening on A to produce a change at B , under the assumption that the intervention is suitably local. The assumption of locality here is relative to the other nodes in the model: we assume that an intervention (quantum instrument) directly effects only the node with which it is directly associated (recall Woodward’s four criteria for an intervention, we ask that the quantum interventions satisfy the same constraints).

The generalisation to multipartite nodes is relatively straightforward. Additionally, a distinction between direct and indirect cause can also be made in a fairly intuitive manner by considering possible interventions on *all* nodes of a network. Thus an instrument can identify a direct causal relation between two nodes, if it remains a difference-maker in the above sense (equation. 5.22), for *all* possible choices of instruments at all other nodes in the network. If there is a choice of instrument at another node that breaks the correlation, then the relation is one of *indirect causation*.

This means that “screening off” in this formalism is also an existentially quantified notion. As long as there exists an instrument that can break the correlation between two other nodes, then this node has the potential to act as a screening off node. This is an important feature of this framework as it means that intermediate nodes *can* screen off two adjacent nodes from one another, as distinct from the QDAGs and GDAGs of Chapter 4.

5.6 Quantum Causal Model

We are now in a position to give a more formal definition of a quantum causal model (QCM). First, recall multiple outgoing edges from a single node represent different physical systems. Each system is associated with a tensor factor: the output space of a node factorises as $A_O = \bigotimes_{e \in \text{OUT}_A} S_e$, where OUT_A is the set of edges departing from A and S_e is associated with an



(a) The causal structure of a Quantum Causal Model is represented by a DAG, where the nodes S, A, B, C represent local spatio-temporal regions. (b) Edges connect input and output spaces for both parent and child nodes. The model is represented by a process matrix that satisfies the *causal quantum Markov condition*, Eq. (5.24). For the example in the picture, $W = T_1^{S_I} \otimes T_2^{S_{O_1} A_I} \otimes T_3^{S_{O_2} B_I} \otimes T_4^{(A_{O_1} B_{O_2}) C_I}$, where C_I is formed by the tensor product of the two input systems

Figure 5.7: Markov quantum causal model.

outgoing edge e . A quantum causal model is then a set of unitaries connecting the output spaces of parents of a node, plus some unobserved noise, to the input space of the child node. Essentially the QCM is a list of quantum channels (CPTP maps), that combine to form a DAG (see figure 5.6).

Definition 5. For a given set of local space-time regions $\mathcal{L} = \{L_j = I_j \otimes O_j\}_{j=1}^n$, a **Markov Quantum Causal Model (MQCM)** is a pair $\langle G, W \rangle$, where

(i) $G = \langle \mathcal{L}, \mathcal{E} \rangle$ is a DAG that has the local space-time regions as vertices;

(ii) to each edge $e \in \mathcal{E}$ is associated a space S_e such that $O_j = \bigotimes_{e \in \text{OUT}_j} S_e$, $j = 1, \dots, n$, where

$$\text{OUT}_j := \{e \in \mathcal{E} | e = (L_j, L_k)\} \quad (5.23)$$

is the set of edges departing from the vertex L_j ;

(iii) W is a process matrix of the form

$$W^{L_1 \dots L_n} = \bigotimes_{j=1}^n T_j^{PS_j I_j} \otimes \mathbb{1}^{O_D}, \quad (5.24)$$

where $O_D := \bigotimes_{k \in \mathcal{D}} O_k$ is the output space of the nodes with no outgoing edges, $\mathcal{D} := \{k \mid \text{OUT}_k = \emptyset\}$; $PS_j := \bigotimes_{e \in \text{IN}_j} S_e$ is the *parent space* associated with node L_j , with

$$\text{IN}_j := \{e \in \mathcal{E} \mid e = (L_k, L_j)\} \quad (5.25)$$

the set of incoming edges to L_j ; and

$$T_j^{PS_j I_j} \geq 0, \quad \text{tr}_{I_j} T_j^{PS_j I_j} = \mathbb{1}^{PS_j}, \quad j = 1, \dots, n. \quad (5.26)$$

5.6.1 Quantum Causal Markov Condition

The possibility of causal discovery via intervention is linked, as in the classical case, to a Markov condition. A process matrix that has the same structure as equation 5.24. will factorise over a DAG G . This is analogous to the requirement that classical joint probability distributions factorise over a classical DAG according to the Causal Markov condition. In the quantum case, we can test the causal structure via interventions (using instruments) and ensure we have the correct factorisation of the process matrix. In a similar manner to the classical case, one can also consider the possibility of latent nodes, unobserved regions where quantum events can take place. A model may not be Markovian according to the above condition if there are such unobserved nodes, but as with the classical case Markovianity is restored by including them in the model. It is in fact always possible, given a causally ordered process matrix, to reconstruct a Markovian graph by introducing latent nodes. This is verified by some previous work on quantum networks (Chiribella et al. (2009)) where the authors showed that any causally ordered process matrix can be represented by combining a sequence of quantum channels with some unobserved/uncontrolled quantum memory. In the language of causal modelling, this is

equivalent to the claim that any set of quantum causal mechanisms (channels), coupled to a set of unknown latent nodes, can be represented as an ordered set of causal mechanisms.

5.7 Causal Discovery using Hilbert-Schmidt basis

Discovering causal structure from scientific data is the primary feature that connects the classical causal modelling formalism to empirical evidence. The quantum causal formalism also affords a similar connection. It is possible to expand CJ matrices and process matrices in a particular choice of basis, known as the Hilbert-Schmidt bases. The advantage of using this representation is that it provides an easy way to check if a given (empirically derived) process matrix contains terms corresponding to particular causal connections. This feature of the process matrix formalism is presented in Costa and Shrapnel (2016) [p 13]. To explain how this works, first consider that one can always re-write a density matrix of n qubits as a set of Pauli matrices:

$$\rho = \frac{1}{2^n} \sum_{i_1, i_2, \dots, i_n}^3 k_{i_1, i_2, \dots, i_n} \sigma_{i_1} \otimes \sigma_{i_2} \otimes \dots \otimes \sigma_{i_n} \quad (5.27)$$

where k_{i_1, i_2, \dots, i_n} are real numbers.

A Hilbert-Schmidt (HS) basis is a set of self-adjoint matrices, where

- (i) σ_0 is associated with the identity, $\text{tr } \sigma_0 = 1$
- (ii) for all other σ_n , $\text{tr } \sigma_n = 0$, and
- (iii) the product of any two $\sigma_\alpha \sigma_\beta = d_X \delta_{\alpha\beta}$.¹⁹ One can expand local maps and process matrices in such HS bases, a move that ultimately serves to simplify causal discovery.

For example, an arbitrary CP map for a node A can be expanded as follows:

$$M^{A_I A_O} = \sum_{\mu\nu=0}^{d^2-1} v_{\mu\nu} \sigma_\mu^{A_I} \otimes \sigma_\nu^{A_O}, \quad (5.28)$$

where $v_{\mu\nu}$ is a real number.

¹⁹Pauli matrices are an example for $d = 2$.

Some basic terminology is convenient to use when trying to distinguish among different kinds of possible graphical connections using this HS decomposition method. Connections are dubbed to be of a particular *type*²⁰:

- (i) The term proportional to identity, $v_{00}\mathbb{1}$, is called *of type 1*.
- (ii) Terms equal to the identity on all subsystems except subsystem X are called *of type X* . For example, the terms $v_{0j}\mathbb{1}^{A_I} \otimes \sigma_j^{A_O}$, for $j > 0$, are of type A_O .
- (iii) Terms equal to the identity on all subsystems except $X_1 \otimes \cdots \otimes X_k$ are called *of type $X_1 \cdots X_k$* . For example, the terms $v_{jl}\sigma_j^{A_I} \otimes \sigma_l^{A_O}$, for $j, l > 0$, are of type $A_I A_O$.
- (iv) The sum of terms of different types X_1, \dots, X_r is called *of type $(\sum_{j=1}^r X_j)$* . For example, a CP map associated with node $A = A_I \otimes A_O$, which decomposes into HS bases according to equation 5.28 above will have terms of type $(1 + A_I + A_O + A_I A_O)$.

As we assume that the channels representing causal mechanisms are trace-preserving (recall equation 5.15), this means that if a plausible graph G contains, for example, a channel from A_O to B_I , then terms of type A_O are excluded as possible terms from the respective process matrix. If the process matrix does in fact contain such terms, one can rule out G as the correct graph.

The basic idea here is that a process matrix can be decomposed using HS bases in order to provide an easy way to check for possible causal connections. For example, if the HS decomposition contains terms of type $A_I B_I$, but a possible graph G does not contain such a common cause connection between nodes A and B , it can be ruled out as a plausible causal structure. The quantum causal discovery methods of Giarmatzi and Costa (2016) utilise such comparative methods.

5.7.1 Quantum fine-tuning

Interestingly, for these quantum causal structures there is also an intuitive notion of fine-tuning. Consider a paradigm example of a *classical* unfaithful model: a binary variable C

²⁰The following itemised list is from (Costa and Shrapnel, 2016)[11], see also Oreshkov et al. (2012) for more examples of such decompositions

that takes the value 1 whenever the outcomes of two fair coins A and B are equal, and is 0 otherwise (Pearl, 2000)[48]. The distribution this arrangement generates is consistent with three causal structures: the three possibilities where each pair of variables is marginally independent but dependent conditioned on the third variable. However, only the correct structure ($A \rightarrow C \leftarrow B$) will be maintained under conditions where we alter the parametrisation of the model - for example by biasing one of the coins.

It is possible to produce a quantum analogue of this kind of unfaithful model. Imagine a collider structure consisting of three nodes $A \rightarrow C \leftarrow B$, where $|\phi^+\rangle|\phi^-\rangle$ casually determines $|0\rangle_C$, and $|\psi^+\rangle|\psi^-\rangle$ determines $|1\rangle_C$.²¹ As with the classical case, if one alters the parametrisation (state) for one of the nodes, it is possible to uncover the fine-tuning and recover the correct causal structure.

There is a more technical definition of Faithfulness that one can demand of these Quantum Causal Models, see Costa and Shrapnel (2016)[11]. Using this definition one can show that under the assumption of Faithfulness, a *unique* causal structure can be discovered for an empirically derived process matrix. Furthermore, the set of *unfaithful* quantum causal models is of measure zero (the proof essentially follows the same reasoning as the classical analogue). As is the case with classical models, this is no longer necessarily the case if one allows for the existence of latent common causes.

5.7.2 Effective and ineffective quantum strategies

The ability to discover the structure of QCM via localised interventions is the key feature linking the methods described in this chapter to interventionist causation. In actual fact, this is to some extent a generalisation of what goes on in quantum laboratories all the time. Physicists frequently use quantum state and process tomographic methods to validate the physical implementation of particular circuit structures: a glance through the supplementary methods of most quantum informational experimental protocols will confirm this fact. Typically, single components of the experimental setup are isolated and tested before combining them together in order to make inferences about the functioning of the whole. Ultimately, this is how they

²¹ $|\phi^\pm\rangle = |00\rangle \pm |11\rangle, |\psi^\pm\rangle = |10\rangle \pm |01\rangle$.

can be confident that certain strategies will be effective in achieving the desired ends.

What can *not* be found in the supplemental methods of quantum physics papers, however, is a general strategy for recovering causal structures from experimentally derived statistics, when the order of events is unknown. Whilst the examples in this thesis are relatively easy to follow, for the more complex circuit structures utilised today, identification of correct causal structure will be less transparent. Having a method for causal discovery in such a setting will likely be a useful addition to the usual computational tools.

It is important to distinguish the methods provided by the quantum causal modelling framework from state and process tomography. Quite simply, tomographic protocols require prior knowledge of the causal structure. In a causal discovery scenario such knowledge is not available. For example, one may be given two sets of measurement results and not know whether they represent two parts of a bipartite state (and thus one should apply state tomography) or two measurements on the same system at different times (and thus one should apply process tomography), or various combinations thereof. The aim of causal discovery is precisely to find out the causal structure from data only, without such *a priori* assumptions.

Now at this point, one might argue that quite obviously in *practical* quantum physics one is *not* given black boxes with lights and levers, and expected to deduce the causal structure from these elements alone. There is typically a large amount of background theoretical understanding of the structure available. However, perfectly isolated experiments where one knows (and can control) *all* the causally relevant degrees of freedom (i.e. one does not have to contend with potentially causally relevant noise) is simply a myth. Initial correlations between a quantum system and its environment (resulting in a latent common cause scenario) are often difficult to detect. Whatever the background theoretical assumptions may be, the reality is often quite different. The methodology of QCM provides a foundation from which to consider “causal tomography” using informationally complete instruments (rather than just informationally complete POVMs) to detect such instances.²²

²²The methods of (Ried et al., 2015) consider such “causal tomography” for two level quantum systems. They do not generalise to other dimensions, nor do they produce a complete framework for quantum causal modelling. This is in part due to the different definitions of “observational” vs “interventionist” data: they consider measurement outcomes to be passive observations, rather than interventions.

Furthermore, the blossoming field of quantum machine learning will likely also benefit from having a clear formalism for recovering quantum causal structure. Probabilistic graphical methods form one of two key approaches in classical machine learning (the other being neural networks), as such the quantum version presented here opens up the possibility for the development of an analogous field.

5.8 The emergence of classical causal structure.

One of the key requirements of the quantum causal modelling framework, as discussed earlier, is that one must be able to recover classical causal models in a suitable limit. Recall, we are not in the business of simply throwing away the classical causal modelling framework and replacing causal explanation with our own quantum version. Rather, we wish to ensure that we can still recover the classical formalism as a limiting case.

The process formalism of QCMs affords two complementary ways of viewing the relationship between classical and quantum causal structure. All quantum causal models can be shown to reduce to classical ones as a limiting case, and conversely, all classical models can be given quantum causal representation with the same structure. One can define “effective” classical causal models: quantum models where it is only possible to access and manipulate states in a fixed basis. That is, the allowed interventions correspond to classical interventions. As long as this basis factorises over separated systems it will be possible to recover the statistics of the respective classical causal model.

Alternatively, one can view classical causal models as a kind of limiting case of quantum causal models, where the choice of instruments are restricted to be diagonal in a particular ‘pointer’ basis. If one is restricted to such a subset of instruments, then the statistics generated will also match those of a classical causal model.

First we consider the situation where we are only able to access states in a fixed basis. We can represent this case by restricting the CP maps realised at each node to be of a particular form. Recall, for quantum causal models we make a distinction between incoming states, outgoing states and measurement *outcomes*. Maintaining this conceptual distinction, first consider a

quantum causal model with incoming states (labelled z_j), outgoing states (labelled o_j), and measurement outcomes (labelled x_j) (see figure 5.8).

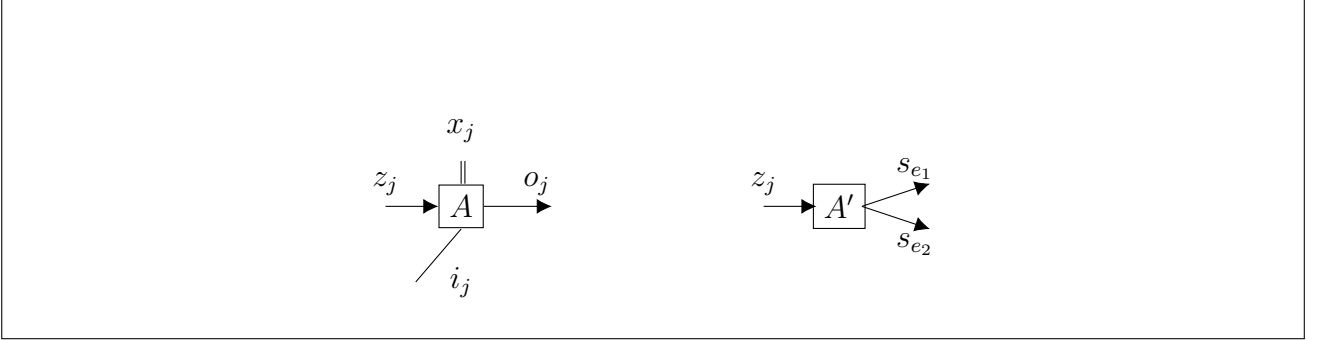


Figure 5.8: For a given basis, z_j labels the possible *state* of the incoming system, o_j labels the outgoing state of the system, x_j labels the possible measurement *outcome* and i_j labels the possible instruments. The figure on the left has a single outgoing system, the figure on the right has two outgoing systems.

In the case where there are multiple outgoing systems (the right hand figure), we assume that the output space associated to the node factorises over the outgoing edges, S_e :

$$O_j = \bigotimes_{e \in \text{OUT}_j} S_e, \quad (5.29)$$

denoting \vec{o}_j as the set of all outgoing edges associated to the node:²³

$$\vec{o}_j := \{s_e\}_{e \in \text{OUT}_j} \quad (5.30)$$

with

$$[\vec{o}_j]^{O_j} := \bigotimes_{e \in \text{OUT}_j} [s_e]^{S_e} \quad (5.31)$$

For classical situations, the incoming (z_j) and outgoing (\vec{o}_j) states are effectively hidden from us: we only have access to the distribution of measurement outcomes, conditioned on the choice of instrument. If we define CP maps of the form:

²³For the remainder of this section, we utilise the notation $[z_j] := |z_j\rangle\langle z_j|$ (Cohen, 2011)

$$\bar{M}_{x_j|i_j}^{I_j O_j} = \sum_{z_j \vec{o}^j} P(\vec{o}^j, x_j | z_j, i_j) [z_j]^{I_j} \otimes [\vec{o}^j]^{O_j}, \quad (5.32)$$

then for a fixed basis we can interpret the CP maps as capturing an incoming classical system in state z_j , undergoing a measurement determined by choice i_j and resulting in an outcome x_j . For each outgoing edge e there is also an outgoing system in state s_e (see figure 5.8). Equation 5.32 reflects the fact that the conditional probability for the measurement outcome and the state of the outgoing system, given the state of the incoming system, is determined by the parameter i_j .

In order to remain consistent with the fact that the incoming and outgoing states (z_j and o_j) are *hidden* in classical situations, we place a further restriction on the possible CP maps: the conditional probability must factorise according to:

$$P(\vec{o}^j, x_j | z_j, i_j) = P(\vec{o}^j | x_j) P(x_j | z_j, i_j). \quad (5.33)$$

We justify the fact that the output state \vec{o}^j is independent of the incoming state z_j by the observation that in the usual construction of classical causal models one can ignore the hidden variable z_j . The output state is thus only *indirectly* dependent on the incoming state, via the measurement event.

The statistics generated by a QCM (G_q) subject to these two constraints (equations 5.32 and 5.33) are equivalent to those generated by a classical causal model (G_c) with isomorphic causal structure. That is, for a QCM $\langle G_q, W \rangle$ consisting of nodes L_j , obtained from CP maps of the form in equation 5.32, with probabilities obeying equation 5.33,

$$P(x_1, \dots, x_n | i_1, \dots, i_n) = \text{tr} \left[\bigotimes_{j=1}^n \bar{M}_{x_j|i_j}^{L_j} \cdot W^{L_1 \dots L_n} \right], \quad (5.34)$$

with this conditional probability obeying the Causal Markov Condition for a DAG G_c iso-

morphic to G_q .²⁴

Recall we identified the fact that the distinction between passive observational data and interventionist data can not be made in the quantum case. We can however, give an understanding of classical non-invasive measurements: such measurements correspond, in the quantum formalism, to projective measurements in a fixed basis. Repeating these kind of measurements confirms previously obtained outcomes, they can thus be interpreted as revealing pre-existing properties of the system without disturbing it. Although this is an interesting approach, it does not, of course, tell us *why* the maps and probabilities should be restricted in such a manner. Other approaches that attempt to explain the link between the quantum and classical worlds may be of use here, for example, decoherence theory (Schlosshauer, 2010) or collapse models (Ghirardi et al., 1985).

An alternative approach is to *start* with a classical model (G_c) and replace the random variables (X_j) with quantum nodes ($L_j = I_j \otimes O_j$), where the input space I_j has a basis element $|x_j\rangle$ for each classical value of X_j . The output space (O_j) is the tensor product of a copy of I_j for *each* outgoing edge, resulting in a process matrix:

$$W = \bigotimes_{j=1}^n T_j^{PS_j I_j}, \quad (5.35)$$

$$T_j^{PS_j I_j} = \sum_{z_j ps_j} P(z_j | ps_j) [ps_j]^{PS_j} \otimes [z_j]^{I_j},$$

The coefficient $P(z_j | ps_j)$ is given by the classical model, i.e. ($X_j = z_j | PA_j = ps_j, I_j = \text{idle}$). If only projective measurements in the pointer basis are performed at each node, then the probability distribution will match the expected classical distribution for undisturbed ‘observational’ data:

$$M_{x_j}^{L_j} = [x_j]^{I_j} \bigotimes_{e \in \text{OUT}_j} [x_j]^{S_e}. \quad (5.36)$$

Interventions in this scenario are thus operations of the form (5.32). An arrow-breaking intervention is realised by ignoring the input (z_j) and preparing the chosen value (x_j) on the

²⁴For the proof, see (Costa and Shrapnel, 2016) page 14.

output space:

$$M_{\text{do}(x_j)}^{L_j} = \mathbf{1}^{I_j} \bigotimes_{e \in \text{OUT}_j} [\text{do}(x_j)]^{S_j} \quad (5.37)$$

.

Separating classical nodes into input and output in this manner is not entirely new. The *single node intervention graphs* (SWIGs) of (Richardson and Robins, 2013) follow a similar procedure, although not with this express purpose of characterising a possible relationship between classical and quantum causal models.²⁵

Thus the picture emerging from this framework is that classical causal structure is a special subset of more general quantum causal structures. One can choose whether to regard classicality as due to a restriction on the possible local processes or on the possible local interventions. The feature of nature that enforces these restrictions will not be decided by this formalism alone, and more work is needed to consider this framework in the light of other, more familiar interpretive stances.

5.9 Bell's objections

Now we have an understanding of the formal structure of the quantum causal models, we can review some philosophical implications. For many, the key question is whether the models do in fact provide a causal explanation for the Bell correlations. Recall the historical trajectory:

- (i) EPR suggest that quantum correlations force a choice between non-locality (causal influence between space-like separated events) and completeness,
- (ii) Bell shows that if causal influence is defined along classical interventionist lines, adding further hidden variables does not circumvent the need for non-locality: that is, one cannot save locality by assuming incompleteness,
- (iii) Wood and Spekkens show that even if one allows for non-local influences, the resulting causal explanation is flawed (due to the presence of fine-tuning): that is, one cannot save causal explanation by allowing for non-local causal influences.

²⁵This is also noted in (Ried et al., 2015).

The quantum causal modelling framework presented here assumes it is the characterisation of causation and causal explanation that is at fault, and seeks to provide an alternative. The alternative presented is complete, in the sense that all correlations can be explained by common cause, direct cause or indirect cause relations. If such an explanation is not possible, then this signals the existence of hidden causes and an extended model that does correctly characterise the empirical statistics can be formed. The formalism recovers the classical modelling formalism as a limiting case, thus we do not need to entirely give up on the interventionist account, with all its intuitive appeal. Rather, it is seen as an approximation of something deeper.

Regarding locality, direct causal influence in these models is consistent with relativistic constraints by virtue of its consistency with quantum mechanics. Thus, empirically derived statistics can always be explained using such models without postulating non-local causal influence (direct causal mechanisms acting between space-like separated regions).

From the perspective of the Bell literature, there are some obvious objections. Recall, Bell was uncomfortable with including a “human” element within an account of causation: correctly capturing possible signalling relations did not seem, for Bell, to be an apt characterisation of causation. We saw in Chapter 1 that this charge has also been levelled at interventionist accounts of causation: whether interventionist causation reduces to agency is a hotly debated question. The interesting point however, is that this problem does not seem to be peculiar to the *quantum* characterisation. It seems to me that one could argue against the need for agency along exactly the same lines as the standard interventionist response (Woodward, 2003). That is, we observe certain regularities in experimental situations (where we have an element of control) that allow us to infer particular causal relations between naturally occurring events (which we do not control). For example, we saw in section 3.3.3 that the explanation for avian magneto-reception characterises the act of measurement as a random encounter between two molecules: there is no requirement of agency in this part of the explanation. The details are inferred from causal knowledge we have gained through controlled experiments. Similarly, one could envisage the measurement outcomes that define the choices of interventions (CPTP maps) for quantum causal models as likewise *naturally* occurring events. What matters is that they

are determined by parameters outside the scope of the model.

The question of how the QCM's relate to Bell's notion of beables is less clear. Certainly, the models are connected to empirical evidence gathered from "the setting of switches and knobs on experimental equipment, the currents in coils and the reading of instruments." However, if Bell's aim was to use beables in order to expunge the use of *any* mathematical representational devices that move beyond classical random variables, then clearly the QCM's fail in this respect.

As I see it, if one believes that interventionism is the correct way to think about causation, then the empirical results of quantum experiments force one of three choices: (i) dub the quantum world as mysteriously acausal, (ii) *abandon* the interventionist account as the preferred account of causation and look for an alternative causal theory that *can* explain quantum correlations, or (iii) *generalise* the interventionist account so that it can account for *both* classical and quantum causal relations. The QCM's of this thesis represent an attempt at the third path.

5.9.1 *Universality*

One of the criticisms of interventionist causation is that one cannot give an account of universal causation. The problem being if one includes the entire universe there is no longer an 'outside' from which to intervene. This issue is equally relevant to the quantum case. If one expands the process matrix to include everything in the universe, then one loses the possibility of outside intervention. Philosophically, this kind of problem is well known: 'going global' often has serious consequences for locally consistent theories that rely on a notion of externality.²⁶ Frisch has presented some counter-arguments to this worry (Frisch, 2014a)[93-100], and it would be interesting to consider whether these arguments would apply equally well for the quantum models presented here.

²⁶The measurement problem is another paradigm example of this.

5.10 Conclusions

It is clear that there is a great deal more that can be said about the quantum causal modelling formalism than I have presented here. What I have shown is that it is possible to have a well defined notion of intervention that can correctly identify quantum causal structure. We also saw that this structure can be used to distinguish between effective and ineffective strategies. As such, this framework ought to be considered as providing an interventionist account of quantum causal structure. A number of parallels between the classical and quantum case were identified (see table 5.1).

There are, of course, some important differences:

- (i) Whilst classical interventions can be associated with clamping a variable to a single value, to take full advantage of a quantum system one needs to utilise more general interventions that determine distributions over sets of events.
- (ii) The mathematical objects used to describe edges and nodes are different. One can, however, represent a classical causal model using the set of quantum mathematical objects, with some added constraints.
- (iii) For the classical case, the fact that unmodelled variables must remain uncorrelated (the independent noise assumption) can be used to justify the use of the Causal Markov Condition. In the quantum case, however, we assume two sources of ‘noise’. The first kind is similar to the classical case: we assume ignorance about the exact state of environmental variables we neglect to include in the model (recall section 3.4). The second kind is different: this is the fundamental randomness that enters via the Heisenberg Uncertainty relations.
- (iv) The quantum causal models presented here are also perspectival in a similar way to their classical counterparts. What we choose to include and what we choose to exclude from our models determines, to some extent, the causal structure of our graphs. In both quantum and classical cases, the consistency across different perspectives is captured by the overall consistency of the modelling formalism. The manner in which different perspectives are

characterised is somewhat different between the two cases however. In the quantum case, the choice of which background variables to include is in general determined by the length of time we require various components of the overall system to remain coherent. It is not clear if there exists a classical counterpart to this requirement.

Classical vs Quantum Causal Models	
Classical Models	Quantum Models
Nodes = collections of classical values	Nodes = collections of CP maps
Surgical Intervention = fixing local value	Surgical Intervention = fixing local CPTP map (e.g. preparing a random state)
Edges = causal mechanisms defining functional relationship of node to parents	Edges = causal mechanisms defining functional relationship of node to parents
Factorisation according to the cartesian product defines causal Markov Condition	Factorisation according to the tensor product defines Quantum Causal Markov Condition
Markovianity recovered by addition of latent nodes	Markovianity recovered by addition of latent nodes
Structures support interventionist inferences	Structures support interventionist inferences
Agent-dependent and independent perspectives equally plausible	Agent-dependent and independent perspectives equally plausible

Table 5.1: Similarities between Classical and Quantum Causal Models.

Conclusions

I have argued in the preceding chapters that an interventionist account of quantum causation is possible.

In the first chapter I argued that interventionist causation is primarily about identifying structures (causal models) that support interventionist queries. That is, the broad aim of causal modelling is to produce models that allow us to determine when one strategy will be effective and another will not. The distinction between the two lies in the definition of an intervention: an action that is determined by a parameter outside the space of the model. It is by virtue of interventions that we are able to identify the manner in which two variables have become correlated, a fact echoed by the slogan *correlation is not causation*.

I also highlighted in Chapter 1 that for Pearl and Woodward localised interventions are possible when systems have a Markovian representation. We saw that for Pearl, the relationship between Markovianity and the possibility of localised intervention was underpinned by his particular interpretational stance. Pearl believes that *deterministic mechanisms* are responsible for causal effects, with probabilities entering solely by virtue of our ignorance of all the facts. Each node in a model is associated with a conditional probability distribution, characterising the effects of incoming causal mechanisms, plus some unobserved noise. A model is Markovian when such noise terms are uncorrelated: correlation can result in latent common causes and alter the structure of the model. For Woodward, we saw that his assumptions are very similar, although the focus is on the notion of intervention, rather than the mathematical details of a particular formalism.

I argued also in Chapter 1 that the assumption of *Faithfulness* is particularly relevant in the context of *causal discovery*. This occurs when one is presented with empirical data, and

attempts to reconstruct a plausible representative causal model. Without Faithfulness, one can trivially satisfy the requirement of Markovianity by ensuring the proposed graph is complete. Additionally, I showed that Faithfulness is also needed to rule out structures that are *fine-tuned*: when a model is fine-tuned the probabilistic independencies are generated for only a *subset* of possible causal parameters.

In the second chapter I argued that the current perspective on interventionism and QM is missing an important point: namely, that one is *not* faced with giving up either Faithfulness or the Markov condition. To do so would negate the value of these assumptions in linking causal structure to interventionist data. Rather, the alternative perspective I advocate *starts* with a definition of a quantum intervention, and develops an appropriate causal account from there. Insofar as it is possible to develop fine-tuned causal models that account for quantum correlations (using retrocausal, superdeterministic and Bohmian models), such models do not respect a basic requirement of interventionist causation: that one can perform interventions to confirm the correct causal structure.

In Chapter 3, I provided motivation for the main task of the thesis. I illustrated that physicists frequently *do* use models that facilitate interventionist inferences. I introduced quantum circuit diagrams and presented three examples from the scientific literature that put such models to work. Lest someone complain that quantum interventionism *only* applies to a certain quantum technologies, I provided a model for a naturally occurring quantum phenomena: the avian magneto-reception model.

In Chapter 4, I considered some alternative quantum causal models. It was shown that none of these frameworks facilitated discovery of causal structure in the interventionist sense presented in this thesis. Whilst Laskey's model included a notion of intervention, discovery was thwarted by her use of two distinct edges (causal and merely correlational). Whilst Pienaar and Brukner's model included a version of d-separation and Markovianity, intervention on intermediate nodes was not possible. As a result, the relationship between classical and quantum causal structure was rendered somewhat ad hoc. For the GDAGs of (Henson et al., 2014), the hybridised structure of the model (consisting of quantum and classical nodes) circumvented the

possibility of a unified notion of intervention and thus a clear method for causal discovery.

In Chapter 5, I presented an account of interventionist causation that *could* be applied to both quantum and classical phenomena. The mathematical objects required to characterise variables and values in this formalism were different to the usual classical versions: sets of CP maps replace values, spatio-temporal regions replace variables, and quantum instruments replace interventions. Using these objects, a new formalism for quantum interventionist causation was presented. We saw that it is possible to use empirical data to build quantum causal structures via this framework, and that these structures reflect possible interventions that physicists can make in order to facilitate a variety of causal inferences. Finally, I presented a number of similarities and differences between the traditional methods of classical interventionism and this newer quantum counterpart.

There are, of course, still many open questions. I have not explored in any depth the perspective that classical causal structure is *emergent* in some sense from quantum causal structure. I do, however, believe that the work of this thesis presents a valuable starting point from which to consider such a claim. Similarly, I have not examined in detail the relationship between quantum causal models and other traditional interpretations of quantum mechanics. Again, I think the quantum modelling formalism presented here may well prove a useful lens through which to consider various causal aspects of such alternatives. This, however, is work for the future.

Bibliography

- Abramsky, Samson, and Bob Coecke. 2009. “Categorical Quantum theory.” In *Handbook of Quantum Logic and Quantum Structures: Quantum logic*, eds. K. Engesser, M. Gabbay Dov, and D. Lehmann. Elsevier.
- Andersen, Holly. 2013. “When to Expect Violations of Faithfulness and Why it Matters.” *Philosophy of Science*. 80(5):672–683.
- . 2014. “A Field Guide to Mechanisms II.” *Philosophy Compass* 9/4:284– 293.
- Bandyopadhyay, J. N., T. Paterek, and D. Kaszlikowski. 2012. “Quantum coherence and sensitivity of avian magneto reception.” *Physical Review Letters* 109(11):110502.
- Baumgartner, M., and I. Drouet. 2013. “Identifying intervention variables.” *European Journal for Philosophy of Science* 3(2):183–205.
- Bell, J. S. 1971. “Introduction to the hidden variable question.” In *Speakable and Unsayable in Quantum Mechanics*, 40–47. Cambridge University Press.
- . 1976. “The theory of local beables.” In *Speakable and Unsayable in Quantum Mechanics*, 57–65. Cambridge University Press.
- . 1990. “La nouvelle cuisine.” In *Speakable and Unsayable in Quantum Mechanics*, 232–248. Cambridge University Press.
- Bell, John S. 1964. “On the Einstein-Podolsky-Rosen paradox.” *Physics* 1(3):195–200.

- Berta, Mario, Stephanie Wehner, and Mark M Wilde. 2016. "Entropic uncertainty and measurement reversibility." *New Journal of Physics* 18(7):073004. URL <http://stacks.iop.org/1367-2630/18/i=7/a=073004>.
- Brown, Harvey R., and Christopher Gordon Timpson. 2014. "Bell on bell's theorem: The changing face of nonlocality." To appear in "50 Years of Bell's Theorem", Mary Bell and Shan Gao (ends.), CUP., URL <http://philsci-archive.pitt.edu/11268/>.
- Butterfield, J. 1992. "David Lewis Meets John Bell." *Philosophy of Science* 59(1).
- Cai, J., G. G. Guerreschi, and H. J. Briegel. 2010. "Quantum control and entanglement in a chemical compass." *Physical Review Letters* 104(22):220502.
- Cai, J., and M.B. Plenio. 2013. "Chemical compass for avian magneto reception as a quantum coherent device." URL <http://arxiv.org/abs/1304.4143>.
- Cartwright, N., and M. Jones. 1991. "How to hunt quantum causes." *Erkenntnis* 35:205–231.
- Cartwright, Nancy. 1979. "Causal laws and effective strategies." *Nous* 13 (4).
- . 2006. "From metaphysics to method: Comments on manipulability and the causal markov condition." *Brit. J. Phil. Sci* 57:197–218.
- Castelvecchi, David. 2015. "Has giant LIGO experiment seen gravitational waves?" *Nature News* 30.09.2015.
- Cavalcanti, E., and R Lal. 2014. "On modifications of reichenbach's principle of common cause in light of bell's theorem." *Journal of Physics A; mathematical and theoretical* 47:424018.
- Caves, Carlton, and Anil. Shaji. 2010. "Quantum-circuit guide to optical interferometry." *Optics Communications*. 283(5):695–712.
- Chaves, R., C. Majenz, and D. Gross. 2015. "Information-theoretic implications of quantum causal structures." *Nature Communications* 6:5766.

-
- Chiribella, G., G. M. D’Ariano, and P. Perinotti. 2009. “Theoretical Framework for Quantum Networks.” *Physical Review A*. 80:022339.
- . 2015. “Quantum from principles.” URL [arxiv:1506.00398v2](https://arxiv.org/abs/1506.00398v2).
- Choi, Man-Duen. 1975. “Positive semidefinite biquadratic forms.” *Linear Algebra and its Applications* 12(2):95 – 100. URL <http://www.sciencedirect.com/science/article/pii/0024379575900580>.
- Cohen, Scott M. 2011. “When a quantum measurement can be implemented locally, and when it cannot.” *Phys. Rev. A* 84 (5):052322.
- Costa, F., and S Shrapnel. 2016. “Quantum causal modelling.” *New Journal of Physics* 18.
- Davies, E., and J. Lewis. 1970. “An operational approach to quantum probability.” *Comm. Math. Phys.* 17:239–260.
- Deutsch, David. 1989. “Quantum computational networks.” *Proceedings of the Royal Society A* 425(1868).
- Dowe, Phil. 1999. *Physical Causation*. Cambridge: Cambridge University Press.
- Dowling J. P., G. J., Milburn. 2003. “Quantum technology: the second quantum revolution.” *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* 361(1809):1655–1674.
- Eberhardt, Frederich. 2007. “Causation and Intervention.” Ph.D. thesis, Carnegie Mellon University.
- Eberhardt, Frederick. 2013. “Direct causes and the trouble with soft interventions.” *Erkenntnis* 79(4):755–777.
- Egg, Matthias, and Michael Esfeld. 2014. “Non-local Common Cause Explanations for EPR.” *European Journal for Philosophy of Science* 4.

- Einstein, A., B. Podolsky, and N. Rosen. 1935. "Can quantum-mechanical description of physical reality be considered complete?" *Phys. Rev.* 47:777–780.
- Evans, P. W. 2014. "Retrocausality at no extra cost." *Synthese* 192(4):1139–1155.
- Evans, Peter W., Huw Price, and Ken B. Wharton. 2013. "New Slant on the EPR-Bell Experiment." *British Journal for the Philosophy of Science* 64:297–324.
- Farr, M., and A. Reutlinger. 2013. "A Relic of a Bygone Age? Causation, Time Symmetry and the Directionality Argument." *Erkenntnis* 78(2):215–235.
- Fisher, Ronald. 1935. *The design of experiments*. Macmillan, 9th edition.
- Fleming, Scholes G. D. Cheng Y.C., G. R. 2011. "Quantum effects in biology." *Procedia Chemistry* 3(1):38–57.
- Frisch, Mathias. 2014a. *Causal Reasoning in Physics*. Cambridge University Press.
- . 2014b. "Physics and the Human Face of Causation." *Topoi* 33:407–419.
- Fritz, T. 2015. "Beyond Bell's Theorem II: Scenarios with arbitrary causal structure." *Communications in Mathematical Physics* online first:1–44.
- Ghirardi, G. C., A. Rimini, and T. Weber. 1985. *Quantum Probability and Applications II: Proceedings of a Workshop held in Heidelberg, West Germany, October 1–5, 1984*, chapter A model for a unified quantum description of macroscopic and microscopic systems, 223–232. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Giarmatzi, Christina, and Fabio Costa. 2016. "A quantum causal discovery algorithm." *In preparation* .
- Glymour, Clark. 2006. "Markov properties and quantum experiments." In *Physical Theory and its Interpretation*, 117–126. Springer.
- Glynn, Luke. 2013. "Causal Foundationalism, Physical Causation and Difference-Making." *Synthese* 190(6):1017–103.

-
- Gutoski, Gus, and John Watrous. 2007. "Toward a general theory of quantum games." *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing* 565–574.
- Hájek, Alan. 2003. "Conditional Probability Is the Very Guide of Life." In *Probability is the Very Guide of Life: The Philosophical Uses of Chance*, eds. Kyburg Jr, E. Henry, and Mariam Thalos, 183–203. Open Court.
- Hardy, Lucien. 2001. "Quantum theory from five reasonable axioms." URL <http://arxiv.org/abs/quant-ph/0101012>.
- Hausman, D., and J. Woodward. 2004a. "Manipulation and the Causal Markov Condition." *Philosophy of Science* 71(5):846–856.
- Hausman, D. M., and J. Woodward. 1999. "Independence, invariance and the causal Markov condition." *Brit. J. Phil. Sci.* 50(4):521–583.
- . 2004b. "Modularity and the Causal Markov Condition: A Restatement." *Brit. J. Phil. Sci.* 55(1):147–161.
- Hausman, Daniel M. 1999. "Lessons From Quantum Mechanics." *Synthese* 121(1-2):79–92.
- Healey, Richard. 2009. "Causation in Quantum Mechanics." In *The Oxford Handbook of Causation*, ed. Hitchcock C. Menzies P. BeeBee, H., 637–686. Oxford University Press.
- Hensen, B., H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenber, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. 2015. "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres." *Nature* 526.
- Henson, Joe, Raymond Lal, and Matthew F Pusey. 2014. "Theory-independent limits on correlations from generalized Bayesian networks." *New Journal of Physics* 16(11):113043.

- Hitchcock, Christopher. 2007. "What Russell got right." In *Causation, Physics, and the Constitution of Reality*, eds. Huw Price, and Richard Corry, 53–54. Oxford: Oxford University Press.
- . 2012. "Events and Times: A Case Study in Means-Ends Metaphysics." *Philosophical Studies* 160(1):79–96.
- Hofer-Szabó, G., M. Rédei, and L.E. Szabó. 2013. *The Principle of the Common Cause*. Cambridge University Press.
- Hofer-Szabó, G., and P. Vecsernyes. 2012b. "Bell inequality and common causal explanation in algebraic quantum field theory." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* .
- Huelga, Susana F, and Martin B Plenio. 2013. "Vibrations, Quanta and Biology." *Contemporary Physics* 54(4):181–207.
- Ismael, Jenann. 2015. "How do causes depend on us? The many faces of perspectivalism." *Synthese* 191(1):245–267.
- . forthcoming. "Against Globalism About Laws." In *Experimentation and the Philosophy of Science*, eds. Bas van Fraassen, and Isabelle Peschard. Chicago: University of Chicago Press.
- Jamiolkowski, A. 1972. "Linear transformations which preserve trace and positive semidefiniteness of operators ." *Reports on Mathematical Physics* 3(4):275 – 278.
- Kochen, S., and E Specker. 1967. "The Problem of Hidden Variables in Quantum Mechanics." *Journal of Mathematics and Mechanics*. 17.
- Koller, Daphne, and Nir Friedman. 2009. *Probabilistic Graphical Models: principles and techniques*. MIT press.
- Korb, Kevin B., and Ann E. Nicholson. 2010. *Bayesian Artificial Intelligence, Second Edition*. Boca Raton, FL, USA: CRC Press, Inc., 2nd edition.

-
- Kost, Matthias, Jianming Cai, and Martin Plenio. 2014. “Resolving Single Molecule Structures with Nitrogen-Vacancy Centers in Diamond.” arXiv:1407.6262.
- Lambert, Chen Y. N. Cheng Y.-C. Li C.-M. Chen G.-Y. Nori F., N. 2012. “Quantum biology.” *Nature Physics* 9(1):10–18.
- Laskey, K. 2007. “Quantum causal networks.” *Proc. AAAI Spring Symp. on Quantum Interaction* 142.
- Leifer, M. S., and R. W. Spekkens. 2013. “Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference.” *Phys. Rev. A* 88(5):052130.
- Maudlin, Tim. 2011. *Quantum non-locality and relativity: Metaphysical intimations of modern physics*. John Wiley & Sons.
- . 2014a. “Reply to werner.” *ArXiv e-prints* arXiv:<https://arxiv.org/abs/1408.1828>.
- . 2014b. “What bell did.” *Journal of Physics A: Mathematical and Theoretical* 47(42):424010.
- Milburn, G. J. 2005. “Foundations of quantum technology.” *Journal of Computational and Theoretical Nanoscience* 2(2):161–179.
- Näger, P. M. 2015. “The causal problem of entanglement.” *Synthese* 1–29. URL <http://dx.doi.org/10.1007/s11229-015-0668-6>.
- Needham, Bradford James Bulpitt Andrew, Chris, and Andrew. Westhead. 2007. “A Primer on Learning in Bayesian networks for computational Biology.” *PLoS Computational Biology*. 3(8).
- Ney, Alyssa. 2009. “Physical Difference Making and Causation.” *British Journal for the Philosophy of Science*. 60:737–764.
- Nielsen, M., and I. Chuang. 2000. *Quantum Computation and Quantum Information*. Cambridge University Press.

- Norsen, Travis. 2015. “Are there really two different bell’s theorems?” *arXiv preprint arXiv:1503.05017* .
- Oeckl, Robert. 2003. “A ”general boundary” formulation for quantum mechanics and quantum gravity.” *Physics Letters B* 575:318–324.
- Oreshkov, O., F. Costa, and C. Brukner. 2012. “Quantum correlations with no causal order.” *Nature Communications* 3:1092.
- Oreshkov, O., and C. Giarmatzi. 2015. “Causal and causally separable processes.” URL <http://arxiv.org/abs/1506.05449v2>.
- Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 1st edition.
- . 2009. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition.
- Pearl, Judea. 2001. “Bayesianism and causality: or, why I am only a half-Bayesian.” 19–36.
- Peres, Asher. 2000. “Classical interventions in quantum systems. I. The measuring process.” *Physical Review A* 61:022116.
- Pfaff, Wolfgang, Bas J Hensen, Hannes Bernien, Suzanne B van Dam, Machiel S Blok, Tim H Taminiau, Marijn J Tiggelman, Raymond N Schouten, Matthew Markham, Daniel J Twitchen, and Ronald Hanson. 2014. “Unconditional Quantum Teleportation Between Distant Solid-State Quantum Bits.” *Science* 345(6196):532–535.
- Pienaar, J., and C. Brukner. 2015. “A graph-separation theorem for quantum causal models.” *New Journal of Physics* 17:073020.
- Price, H. 2007. “Causal perspectivalism.” In *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, eds. H. Price, and R. Corry, 250–292. Oxford University Press.

-
- Price, Huw. 2013. “Causation, Intervention and Agency - Woodward on Menzies and Price.” In *Making a Difference.*, eds. Hitchcock Christopher Beebe, Helen, and Huw Price. Oxford University Press.
- Price, Huw, and Peter Menzies. 1993. “Causation as a secondary quality.” *Journal for the Philosophy of Science* 44:187 – 203.
- Prugovecki, E. 1977. “Information-Theoretical Aspects of Quantum Measurement.” *International Journal of Theoretical Physics* 16(5):321 – 331.
- Rajapakse, J. C., and J. Zhou. 2007. “Learning effective brain connectivity with dynamic Bayesian networks.” *Neuroimaging* 37(3):749–60.
- Redhead, Michael. 1987. *Incompleteness, Nonlocality and Realism*. Oxford Press.
- Reichenbach, H. 1956. *The Direction of Time*. University of Los Angeles Press.
- Richardson, T. S., and J. M. Robins. 2013. “Single World Intervention Graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality.” *CSSS Working Paper* .
- Ried, Katja, Megan Agnew, Lydia Vermeyden, Dominik Janzing, Robert W Spekkens, and Kevin J. Resch. 2015. “A quantum advantage for inferring causal structure.” *Nature Physics* 11:1745–2473.
- Robledo, Lucio, Lilian Childress, Hannes Bernien, Bas Hensen, Paul F A Alkemade, and Ronald Hanson. 2011. “High-Fidelity Projective Read-Out of a Solid-State Spin Quantum Register.” *Nature* 477(7366):574–578.
- Rodgers, Hore P. J., C. T. 2009. “Chemical magneto reception in birds: the radical pair mechanism.” *Proceedings of the National Academy of Sciences* 106(2):353–360.
- Russell, Bertrand. 1912. “On the Notion of Cause.” *Proceedings of the Aristotelian Society* 13:1–26.

- Schlosshauer, Maximilian A. 2010. *Decoherence and the Quantum-To-Classical Transition*. Berlin: Springer.
- Schumacher, B. 1996. "Sending entanglement through noisy quantum channels." *Phys. Rev. A* 54:2614–2628.
- Shrapnel, Sally. 2014. "Quantum Causal Explanation: Or, Why Birds Fly South." *European Journal for Philosophy of Science* 4:409–423.
- . 2015. "Discovering Quantum Causal models." URL <http://philsci-archive.pitt.edu/11829/>.
- Skyrms, Brian. 1984. "EPR: Lessons for Metaphysics." *Midwest Studies in Philosophy* 9.
- Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*. Cambridge, Mass.: MIT Press.
- Steel, D. 2006. "Comment On Hausman & Woodward On The Causal Markov Condition." *Brit. J. Phil. Sci.* 57(1):219–231.
- Stinespring, W.F. 1955. "Positive functions on c^* -algebras." *Proc. Am. Math. Soc.* 6:211–211.
- Strevens, M. 2008. *Depth: An Account of Scientific Explanation*. Harvard University Press.
- Suárez, Mauricio, and Iñaki San Pedro. 2011. "Causal Markov, Robustness and the Quantum Correlations." In *Probabilities, Causes and Propensities in Physics*, ed. Mauricio Suárez, 173–193. Dordrecht: Springer.
- 't Hooft, Gerard. 2009. "Entangled quantum states in a local deterministic theory." arXiv:arxiv:0908.3408.
- Tal, Eran. 2013. "Old and new problems in philosophy of measurement." *Philosophy Compass* 8(12):1159–1173.

-
- Tiersch, M., and H. J. Briegel. 2012. “Decoherence in the chemical compass: the role of decoherence for avian magneto reception.” *Philosophical transactions of the royal society A: mathematical, physical and Engineering Sciences* 370.
- Timpson, Christopher. 2013. *Quantum Information Theory and the Foundations of Quantum Mechanics*. Cambridge University Press.
- Tucci, R. R. 1995. “Quantum Bayesian Nets.” *International Journal of Modern Physics B* 9:295–337. [arXiv:quant-ph/9706039](https://arxiv.org/abs/quant-ph/9706039).
- Verma, T., and J. Pearl. 1988. “Causal networks; semantics and expressiveness.” *Proc. Fourth Conf Uncertainty in Artificial Intelligence* 352–359.
- Werner, Reinhard F. 2014a. “Comment on ‘What Bell did’.” *Journal of Physics A: Mathematical and Theoretical* 47(42):424011.
- . 2014b. “What Maudlin replied to.” *arXiv preprint arXiv:1411.2120* .
- Wiseman, Howard, and Gerard. Milburn. 2014. *Quantum Measurement and Control*. Cambridge: Cambridge University Press.
- Wiseman, Howard M. 2014. “The two bell’s theorems of john bell.” *Math. Theor.* 47(arXiv:1402.0351):424001.
- Wiseman, Howard M, and Eric G Cavalcanti. 2015. “Causarum investigatio and the two bell’s theorems of john bell.” *arXiv preprint arXiv:1503.06413* .
- Wiseman, Howard M, and Eleanor G Rieffel. 2015. “Reply to norsen’s paper” are there really two different bell’s theorems?.” *arXiv preprint arXiv:1503.06978* .
- Wood, C. J., and R. W. Spekkens. 2015. “The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning.” *New Journal of Physics* 17(3):033002.
- Wood, Christopher J., Jacob D. Biamonte, and David G. Cory. 2015. “Tensor networks and graphical calculus for open quantum systems.” *Quantum Info. Comput.* 15(9-10):759–811.

- Woodward, James. 2003. *Making Things Happen. A Theory of Causal Explanation*. Oxford: Oxford University Press.
- . 2007. “Causation with a Human Face.” In *Causation, Physics, and the Constitution of Reality*, eds. Huw Price, and Richard Corry, 66–105. Oxford: Oxford University Press.
- . 2012. “A Functional Account of Causation.” arXiv:PhilSci archive.
- . 2015. “Methodology, Ontology, and Interventionism.” *Synthese* 192(11):3577–3599.
- . 2015a. “The Problem of Variable Choice.” *Synthese* 1–26. URL <http://dx.doi.org/10.1007/s11229-015-0810-5>.
- . 2015b. “Interventionism and the Missing Metaphysics: A Dialogue.” In *Metaphysics and the Philosophy of Science*, eds. Slater, and Yudell. Oxford University Press.
- Zhang, Jiji, and Peter Spirtes. 2015. “The Three Faces of Faithfulness.” *Synthese*. 1:17.

Appendices

Appendix A: Quantum Causal Explanation: or, why birds fly south

The following publication reports work completed during the first year of PhD candidature. This work demonstrates that certain quantum explanations can be given an interventionist interpretation. It is therefore consistent with the exposition of this thesis. However, it is included here as an appendix because the primary aim of this paper was to show that causal explanation was possible, rather than to explicitly characterise an interventionist quantum causal modelling formalism. The paper has been reproduced below for convenience, and in accordance with the University of Queensland requirements, in the same format as the thesis. The original published article can be found here: <http://link.springer.com/article/10.1007/s13194-014-0094-5>

Quantum causal explanation: or, why birds fly south

Sally Shrapnel (received: 27 December 2013 / Accepted: 11 June 2014).

Abstract

It is widely held that it is difficult, if not impossible, to apply causal theory to the domain of quantum mechanics. However, there are several recent scientific explanations that appeal crucially to quantum processes, and which are most naturally construed as causal explanations. They come from two relatively new fields: quantum biology and quantum technology. We focus on two examples, the explanation for the optical interferometer LIGO and the explanation for the avian magneto-compass. We analyse the explanation for the avian magneto-compass from the perspective of Woodward's interventionist theory and provide a causal model. Furthermore, we show how worries expressed by Woodward about quantum causation are circumvented in these cases, concluding that these kinds of explanations are most naturally construed as causal.

1 Introduction

There are several recent scientific explanations that appeal crucially to quantum processes. These explanations come from two relatively new fields, from so-called quantum biology, and from quantum technology (also known as Quantum Engineering). The explanations from these fields involve a concept known as "coherent quantum control" and have a distinctly mechanistic feel. *Prima facie*, these explanations are causal.

Many philosophers believe that scientific explanations display an asymmetry derived from underlying causal structure (Craver 2009; Salmon 1971; Strevens 2008). From this perspective, the explication of the phenomena listed above provides clear support for the existence of quantum causation. The philosophical literature, however, contains considerable skepticism concerning objective quantum causal structure (Hausman and Woodward 1999; Van Fraassen 1982; Woodward 2007). Typically, those who eschew quantum causation derive their arguments from analysis of the correlations that appear in EPR type experiments, the problem being that

whilst the measurement results in such experiments are clearly strongly correlated, they are thought not to be so in virtue of underlying causal structure. Suarez has been a minority voice challenging this orthodoxy (Suarez 2007; 2013) see also (Wood and Spekkens 2012). Other philosophers have an even larger target in their sights, and insist that physics itself is not a domain that admits our causal intuitions with any clarity (Norton 2007; Woodward 2007). Frisch has recently provided some convincing counterarguments to this “neo-Russellian” perspective (Frisch 2012), and the arguments of this paper provide further support for his position.

There are several features of quantum theory that conspire to make the quantum domain a particularly difficult one to analyse in terms of causal structure. The non-local correlative properties of entanglement, the so-called “measurement problem”, and the dispute concerning the ontic or epistemic status of the wave-function all make quantum causation a tricky topic to approach. In this paper we will not address these well-known problems directly. Rather, we show that current scientific explanatory practice suggests that quantum causation is possible. Furthermore, analysis of these explanations in the framework of a standard philosophical approach to causal-explanation, Woodward’s Interventionist approach, supports the possibility of quantum causation.¹ Since Woodward himself feels there are problems for quantum causation, this paper will also go some way towards addressing his concerns.

It is worth noting here that we will not be attempting to give a metaphysical story about wave-functions, superposition states or entanglement. Woodward’s causal theory is not a metaphysically motivated theory and as such we can use the framework to identify causal relations and relations without needing to commit to a particular metaphysical interpretation. Considering the variety of metaphysical interpretations of quantum mechanics currently on offer, we consider this feature of the interventionist approach a virtue.

In Section 1 we discuss some recent examples from the scientific literature that place quantum mechanisms within complex explanatory stories. We will first look at the explanation for one of the novel metrological tasks that is now being attempted through the explicit control and manipulation of a quantum system, and then at an example from the field of ‘quantum biology’. The first example will be used to illuminate an important concept, coherent quantum control,

and the second to ground our discussion of quantum causal explanatory models. In Section 2 we will give a brief overview of a contemporary causal-explanatory approach, Woodward's Interventionist Theory, and discuss some of Woodward's concerns with respect to quantum causation. In Section 3 we will provide causal models for the second example of Section 1. They are represented using the directed acyclic graph format formulated by Pearl, with structural equations based on Hamiltonian quantum mechanics.²⁷

2 Quantum causal explanations

There has been an interesting development in scientific explanation involving quantum mechanics over the last few years. Control and manipulation of quantum coherence in single quantum systems has become a reality in the lab environment (Dowling and Milburn 2003; Milburn 2005) and some intriguing discoveries in biology have prompted claims that perhaps nature has already managed to do some quantum control engineering of her own (Lambert et al. 2012). The claims from quantum biology are not above controversy, but are increasingly being accepted by many scientists across a wide range of fields, and have also inspired some serious attempts at reverse engineering new technology (Cai, Guerreschi, and Briegel 2010). It is consequently not our aim to discuss the validity of the claims made in this new field, but rather to examine the fact that these theories seem to claim quantum states are capable of assuming various difference making roles in multi-level explanations and are hence playing an indispensable part in complex scientific explanation.

Quantum biology is an emerging field that investigates the dynamic quantum mechanisms that have been implicated in several explanatory stories involving natural systems. Perhaps the two most thorough accounts are those of the avian magneto-compass (Rodgers and Hore 2009) and the photosynthetic light-harvesting complex (Fleming, Scholes, and Cheng 2011). In both cases the dynamics of quantum states are thought to provide mechanisms for explaining

²⁷Whilst we use Woodward's interventionist theory to support the conclusion that these explanations are causal, we are not committed to this being the only applicable causal theory. As an example, it is likely that Dowe's Conserved Quantity theory will also be applicable in this context; the details of energetic transfer suggested by the Hamiltonian mechanical description of the system dynamics will likely pick out similar causal relations and relations.

effects that cannot be explained in classical terms alone. The relatively new field of quantum technology also provides us with an opportunity to examine scientific quantum explanations that seem to appeal explicitly to causal structure. This field explores new kinds of technology that we can now develop by harnessing quantum effects to achieve new ends.

In the examples from both quantum biology and quantum technology, the quantum effects referred to are postulated to play a functional and dynamical role, and this can be articulated in terms of manipulation and control of quantum coherence. It is this new development that provides opportunities for applying our usual philosophical tools for causal identification, such as Woodward's manipulationist approach. We do not dispute that there can be other kinds of quantum explanations that do not involve causation, rather that our particular examples appear to demand quantum causal explanation.

It is possible that the explanations selected here are particularly amenable to being cast in a causal framework (specifically the interventionist framework) due to their multi-level nature. However, it is beyond the scope of this paper to address the specific issues surrounding multi-level explanation. Whilst the avian magneto-compass in particular provides a remarkable example of such an explanation, spanning many orders of magnitude and utilising the theories of multiple scientific domains, we will not have anything distinctive to say on the matter here.

It will also become clear that the causal structures discussed in this paper involve what could be considered both "downwards" and "upwards" causation. We later see that our causal structural models include both classical causes that are related to quantum effects, and quantum causes that are related to classical effects. The well known difficulties of inter-level causation and the causal exclusion argument will also not be addressed in this paper, though see (Campbell and Bickhard 2011; Kim 1999; Woodward 2008) for further discussion.

At this point it is perhaps worth pausing to consider what has changed in the quantum physics community to allow quantum states to now enter into such complex, multi-level explanations. It is likely that there are many contributing factors, but possibly the most relevant breakthrough of the last few decades is physicists' ability to isolate and probe single quantum systems. The experimental advances afforded by the work of David Wineland, Serge Haroche

and colleagues have enabled the precise control and manipulation of single quantum systems, and recently earned them the 2012 Nobel Prize for Physics.

By improving our ability to isolate quantum systems, coherent quantum states in the lab can now be maintained for longer time periods, and this means that the mathematical models predicting quantum interactions and coherence times for particular set ups can be made more precise. Thus for a number of situations we can now give quantitative and relatively precise time estimates for the maintenance or loss of quantum effects such as coherence or entanglement. For such systems we can consider that a close match between theoretical prediction and experimental outcome validates that our explanatory model is successfully taking into account all the degrees of freedom, and the details of their interactions, that are relevant for our target explanandum (Dowling and Milburn 2003; Milburn 2005).

An intuitive feel for quantum coherence can be given by appealing to our first example, the explanation for the operation of the gravitational wave detector, LIGO. LIGO is an enormous optical interferometer that has been designed to measure the effects of gravitational waves by monitoring the relative displacement of two mirrors. The measurement sensitivity of this device is such that we can now use it to measure displacements of around 1 part in 10^{18} , an astounding feat that relies crucially on maintenance of quantum coherence (B. Abbott, Abbott, Adhikari, and Ajith 2009).

The explanation can be built up by first considering a single beam splitter device. Photons from an upper source are fired at a semi-silvered mirror (beam splitter) and subsequently detected at both upper and lower photon detectors. For each input photon there is only ever a single detection event; we never detect two photons simultaneously arriving at the output detectors. Multiple trials will result in a probability distribution of 0.5 at either detector; half of the time a photon arrives at the upper detector and half the time a photon arrives at the lower detector. Our usual classical thinking leads us to infer that each individual photon is either being reflected or transmitted at the beam splitter, with random probability for each single trial.

Consider the original setup extended to contain two successive beam splitters, forming a

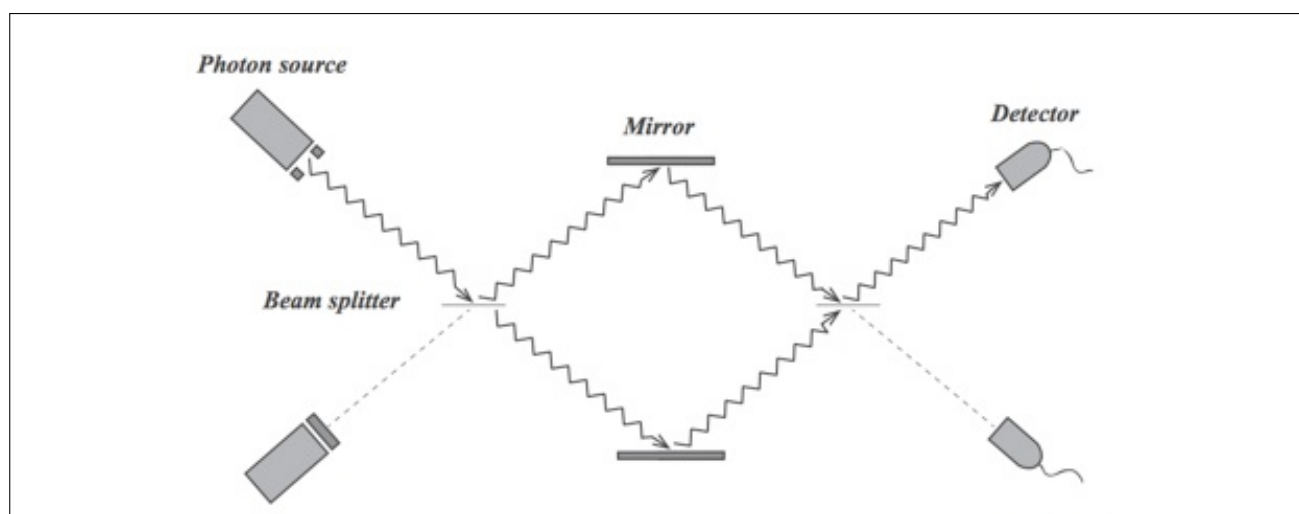


Figure 5.9: Mach-Zender Interferometer

device known as a Mach-Zender interferometer (see Fig. 5.9 below). The addition of the two perfectly reflecting mirrors acts to redirect the photons back to the next beam splitter. Given the inferences from the setup with the single beam splitter (single photons are either reflected or transmitted at each beam splitter), our classical intuition leads us to expect that the outcome probability distribution, for a series of trials, will again be 0.5 for both upper and lower detection events. However, this is not in fact what we see. The outcome of the final detection event is rendered certain by the addition of the second beam splitter; all detection events now occur at the upper detector. The relationship between the input and output values in this system allows us to claim that the system is displaying quantum coherence. Validation of this claim can be made by varying the probability distribution of the input events (firing photons from a range of both upper and lower sources, say 40 % of the time from the upper and 60 % of the time from the lower), and showing that this results in a corresponding change in the probability distribution of the output events.

Formally, this relationship between input and output distributions is captured by both the linearity of Hilbert space (allowing for superposition states) and the Born rule (allowing calculation of the output probabilities). Thus, the ‘machinery’ of what occurs in between input and output events is captured at the level of probability amplitudes. An essential difficulty with quantum mechanics concerns what probability amplitudes actually represent, at least in terms of our usual notion of the kind of variables we meet in the classical world. Again, for the

purposes of this paper, we leave these metaphysical issues aside. We will later see that they do not preclude us from identifying causal relations and relations in the usual way.

By introducing one more element into the picture, we can give both a description of the interferometer LIGO, and also get an intuitive feel for quantum coherence. Consider now that there is an external source of random noise that effects the position of the top- most perfectly reflecting mirror. This random noise can degrade the correlation between the probability distributions of input and output events. Remember we were previously able to track the coherence of the system by varying the probability distribution of the input events and seeing the correlation with the variation in the probability distribution of output events. This correlation can be predicted accurately by numerically tracing the changes in probability amplitudes for the quantum state of a photon as it passes through the device. It is this loss of ability to track coherences, in this case due to the random wobble of the mirror, that is known as decoherence.

Now consider an alternative scenario. Instead of introducing a random wobble in the position of the uppermost mirror, let us now change its position to a fixed new position by moving it northwards (i.e. up the page). If the mirror remains fixed in this new position, there will no longer be detection at the upper detector with certainty, but a proportion of detection events will occur at the lower detector. The important difference this time comes when we vary the probability distribution of input events. Once again we will have perfect correlation between the variation in probability distribution of input events with the distribution of output events, and the system is back to displaying quantum coherence. The “control” part of the term coherent quantum control comes from this ability to accurately determine the probability distribution of output events, for a range of values of both input variables and mirror positions. LIGO is actually a huge folded optical interferometer that takes advantage of this very phenomenon. With arms spanning around 4 km, a tiny repeated wobble is introduced into the length of each arm as gravitational waves pass perpendicularly through the device. This wobble periodically alters the relative positions of the two perfectly reflecting mirrors and thus coherently changes the probability distribution of the output detection events with respect to the probability distribution of the input events. Coherence of the system is maintained by the fact that the

period of the gravitational wave is constant. Maintenance of the system coherence is crucial to the sensitivity of the device. Any additional sources of disturbance, be they unknown or simply uncontrollable, can potentially result in decoherence that will quickly degrade the sensitivity of the device.

We have not listed all the potential sources of decoherence that must be accounted for such that LIGO can achieve maximum sensitivity, and as such, the explanation is somewhat idealised. Additionally, we have not discussed other features that ultimately limit the maximum sensitivity of the device, such as shot noise and radiation pressure, and the techniques for minimising these effects. For the purposes of this paper however, they do not present any complications that threaten the main arguments of the paper, but see (B. Abbott, Abbott, Adhikari, and Ajith 2009) for further details.

The current favoured explanation for the remarkable navigational abilities of the European Robin similarly relies on this notion of coherent quantum control. Many birds are able to navigate accurately for thousands of miles each year as they head south for the winter. Current science tells us they do so by detecting the extremely weak Earth's magnetic field lines, and the mechanism responsible is thought to rely crucially on the maintenance of quantum coherence.

The proposed mechanism is known as the radical pair mechanism, and is believed to take place in the retina of the bird. Within a molecule known as cryptochrome, the quantum spins of two electrons are initially entangled in a spin singlet state. One electron is then excited from the ground state by the absorption of an incoming photon, inducing a conformational change in the surrounding molecule and allowing the electrons to spatially separate. This separation then results in inter-conversion of the quantum state of the electron pair between a triplet (spins aligned) and singlet state (spins anti-aligned). This interconversion is considered a fully coherent, reversible quantum dynamical process, occurring due to coupling of the electron spins with separate individual nuclear environments. The dynamics of the inter-conversion are then further modulated by the earth's magnetic field.

The final step involves the electron pair encountering another reactant molecule. Depending on the joint quantum spin state of the electron pair at the time they encounter the reactant

molecules, different chemical products result, and at different rates. These resultant chemicals are assumed to provide a neural signal that piggybacks onto the visual system of the bird. The speculation being that as the bird changes the orientation of its head with respect to the earth's field lines, the consequent modulation of the chemical signalling may result in a change in the distribution of light and dark patches in the bird's field of view.

The most recent behavioural experiments to support the model involve the application of weak, oscillating radio-frequency pulses to disrupt the radical pair mechanism, and consequently the birds' sense direction. These pulses are tuned to the transition frequency between triplet and singlet states and thus disrupt the usual relationship between the joint spin state of the electron pair and the direction of the bird's head. Further studies of this nature are planned for the coming migratory season, with the ultimate aim being explicit quantum control of the birds' direction sense, via control of the spin states of the electron pair. More details can be found in (Bandyopadhyay, Paterek, and Kaszlikowski 2012; Rodgers and Hore 2009); Cai et al. 2010; Cai and Plenio 2013; Tiersch and Briegel 2012). It is worth commenting that there are alternative possible explanations for the birds' remarkable navigational abilities, most notably magnetite based mechanisms. These mechanisms, however, are unable to account for the fact that (i) the compass appears to be an inclination, rather than polarity compass, (ii) the mechanism appears to be light triggered, and (iii) the birds direction sense is able to be disrupted by weak, oscillating radio-frequency pulses. For these reasons the radical pair mechanism has gained support as the favoured current explanation (Lambert et al. 2012). This means that at the very minimum, this paper shows that a causal understanding is possible if the radical pair model is ultimately shown to be the correct one.

As Cai and Plenio (2013) recently pointed out, the explanation for the magneto- compass can be considered analogous to an interferometer (such as LIGO). For the magneto-compass it is the hyperfine coupling between the nuclear environment and electron spins that provides the process equivalent to passage through the beam splitters; formally, it is this coupling that sets up a superposition of states. In the interferometer LIGO, interference occurs between the two paths through the interferometer, whereas in the magneto-compass it occurs in the energy

eigenbasis of the electron-nuclear coupling Hamiltonian. The input and output states of the interferometer LIGO are the mutually exclusive photon paths; the photon is fired with certainty from either the upper or lower source, and is detected with certainty at either the upper or lower detector. In the case of the magneto-compass, the input state is a singlet state, and the detection states can be either one of the mutually exclusive states: triplet or singlet.

The final measurement step in the interferometer occurs when a photon is detected at the output; analogously, the measurement step in the avian magneto-compass occurs when the reactant molecules randomly encounter the radical pair. The changes in the probability distribution for output photons over many trials in the LIGO interferometer allow calculation of the minute changes of the mirror position due to passage of gravitational waves; the relative ratios of singlet and triplet products, averaged over many molecules in the bird's retina allows detection of the direction of the Earth's extremely weak magnetic field. Thus the bird is able to control the relative ratio of singlet to triplet product, and thus the associated visual pattern, by altering the direction of its head with respect to the Earth's magnetic field lines. During orientation, the birds are observed to perform scanning behaviour, repeatedly move their heads before fixing on a direction.

The dynamics of the avian model are expressed using quantum Hamiltonian dynamics, to keep track of the coherences within the system. As discussed above, detailed knowledge of coherent states can be maintained when we have sufficient knowledge and control of any degrees of freedom that influence the system of interest. In general, biological systems are complicated, wet and warm environments and were originally thought to provide too many sources of decoherence for quantum effects, such as interference, to play any functional role. However, recent improvements in the sensitivity of experimental tools means we can now detect changes at extremely short length and time scales. Recent experiments have verified that functional quantum effects (such as coherence and entanglement) can occur at very fast timescales, and are sufficiently well localised such that their effects can make a difference before slower thermal environmental effects can interfere. Furthermore, there is increasing evidence that there is often constructive interplay between thermal noise in stationary non-equilibrium systems, such that

quantum correlations are not suppressed but rather enhanced or regenerated by interaction with the environment (Huelga and Plenio 2013).

It is critical to realise that support for the avian model has come from performing various interventions, and as the interventions have become more sophisticated the model has gained strength and support from the scientific community. There are many kinds of explanations to be found in science, but it is those that support interventions that are often considered most validated. Following the experimental verification by Alain Aspect, the violation of Bell's theorem allowed by quantum theory achieved greater acceptance as really telling us something about the nature of the world. Intervening on the radical pair mechanism of the avian magneto-compass has similarly validated the causal structure of this explanation. Scientists have considered changes in (i) the nuclei surrounding the separated electron pair (spin state and spatial configuration), (ii) the strength or direction of the external magnetic field, and (iii) the joint spin state of the electron pair.

Whether these interventions can be considered as Woodwardian is considered in section 4 below, but first a brief recap of Woodward's interventionist theory of causal explanation.

3 Interventionist causation

Woodward's interventionist theory is motivated by the intuition that we typically associate causal relations with an ability to manipulate or control the value of a particular variable. To identify a causal relation between A and B , given some particular background conditions Z , we wish to both isolate A and also vary its value. Furthermore, we wish to hold the background conditions Z as fixed as possible to eliminate confounding factors, such that we can observe and measure the direct effect A has on B . In this way, we are able to deduce the causal relation by seeing how B changes when we vary A in some way. This intuition is used in the design of randomised, double-blind, placebo controlled medical trials, that in effect allow us to hold as many background factors as possible fixed, whilst assessing a treatment effect. In fact, scientific experiments are often designed to establish causal relationships by holding various factors fixed, whilst varying others and then analysing the consequences, and the fact

that this causal theory closely mirrors the scientific method has meant it has found favour with philosophers of science from many different disciplines. The potential for such manipulations allows us to differentiate between relationships that merely express correlation and those that identify causation (Woodward 2005).

The interventionist account draws on previous work done in causal modelling by (Pearl 2009; Spirtes, Glymour, and Scheines 2000), who discuss causation from the perspective of various causal discovery algorithms. These accounts all require the careful explication of an ideal intervention. Formally, a causal relation between a variable (A) and a variable (B) holds if and only if there exists an ideal intervention (I) on A , such that the value of B changes.

The ideal intervention (I) must meet the following requirements (Woodward 2007):

- (i) I must be the only cause of A - the intervention must completely disrupt the causal relationship between A and its previous causes, so that the value of A is set entirely by I .
- (ii) I must not directly cause B via another route
- (iii) I should not itself be caused by any cause that affects B via a route that does not go through A
- (iv) I leaves the values taken by any causes of B unchanged, except those that are on the directed path from I to A to B , (should this exist).

Requirement 1 ensures the value of A is not effected by any other causal variable Z , but entirely set by I . Requirements 2 and 3 remove any other ways A and B may be correlated other than A 's causing B .

The causal effect here is relativised to a background context Z (which will incorporate information about other causes of B). Thus this account attempts to answer the question "in a given context, what is the difference made to B by varying A ?". In deterministic contexts we need to hold Z steady, in indeterministic contexts we can define causal effects in terms of expectations with respect to probability distributions. The relata in Woodward's causal claims

are called “variables”, where a variable is simply a property, quantity etc., which is capable of taking two or more “values”.

The structural equations framework of Pearl (2009) relativises the causal structure of a system to a causal model, defined as an ordered pair $\langle V, E \rangle$, where V is a set of variables and E a set of structural equations. The structural equations state the relations among variables (deterministically or probabilistically) in the system of interest. The possible states of the system are described by the variables in V , and the structural equations model the dynamical evolution of the system. Each variable in the model has a corresponding structural equation, for exogenous variables the form is $Y = y$ (set at the actual value) and for endogenous variables the form is $Y = f(X_1, X_2, X_3 \dots X_n)$. Such equations are read from right to left, so X is a “parent” of Y . The structural equations are frequently represented as directed acyclic graphs, with the nodes of the graph being the variables in V and the arrows between them representing the causal dependencies between variables.

Woodward quite rightly points out that typically in explanations we wish to know more about our causal models. We don’t only wish to know which variables are causally related, but “... we wish to know which kind of interventions on X have an effect on Y , how they effect Y and under what background circumstances.” (Woodward 2007). This information is captured by two further requirements: invariance and stability. A causal model is invariant if it holds under at least some range of values of the intervention. A model is stable if it holds for at least some range of changes in background variables, where these variables are typically not included in the model.

Woodward has written surprisingly little about applications of his theory to the domain of quantum mechanics. Essentially, he regards the correlations between EPR pairs as providing evidence that the relationship between them is not causal in nature. The defence for this position rests on the assumption that it makes no sense to make an intervention on the joint state of an entangled system, mainly by virtue of the fact that it is not a factorisable state (Hausman and Woodward 1999). Suarez has some counter-arguments to this position, and both views will be discussed further below.

Woodward also has concerns about application of his theory to quantum mechanics considered as a fundamental, universal theory. If the state of the universe can be assigned a quantum state, then it makes no sense to envision an intervention on this. Woodward wishes his theory to apply to what Pearl calls “small worlds” (Woodward 2007). As is discussed below, it is clear that the model for the magneto-compass clearly encapsulates such an isolated system, which can support a notion of external intervention.

4 Causal model for avian magneto-reception

The inclusive causal model will contain many relata. For the purposes of this paper we include only that part of the overall explanation that requires the inclusion of quantum states. For simplicity, we will assume that it is the singlet reaction product that provides the neural signal that ultimately determines the direction sensing ability of the bird. By considering various experiments that have been performed to better understand the radical pair mechanism, it is possible to suggest a simple causal model for the magneto-compass.

There are five variables to consider (see Fig. 5.10), with two of the variables being quantum states. The immediate question that arises in this context is how one can assign a “value” to a quantum state. Although abstract, one can think of the value of a quantum state as being the corresponding ray in Hilbert space, with any changes in value being reflected by various transformations on this ray. Thus the structural equations that relate the variables in this model need to tell us how transformations on one quantum state variable are related to transformations on other quantum state variables in the model. In general, it is possible to predict how various quantum states change by considering the possible energy configurations of the total system, in this case determined by the Hamiltonian for the hyperfine couplings of the system (see Cai and Plenio 2013 for more details).

Variables:

- (i) B : the strength and direction of the external magnetic field. This includes the angle between the direction of the bird’s head and the local magnetic field (Earth or applied) and the strength of the local field. This variable could be further divided into two parent

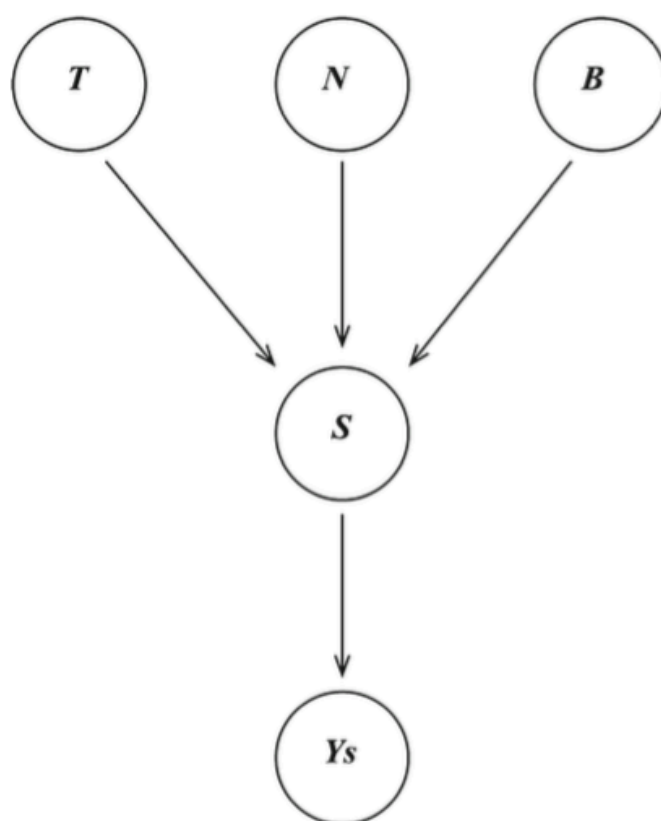


Figure 5.10: Causal diagram for radical pair mechanism.

variables, one variable for the field strength and one for direction, and in general the experiments reflect this separation. For simplicity here we have subsumed this information into one variable, with the possible values of B thus being a function of two parent variables: field strength and direction.

- (ii) N : the nuclear spin state. The nuclei are assumed to be in a thermalised spin state; that is, we consider the spins will flip randomly at a certain frequency. As the electron-nuclear hyperfine dynamics occur on a much faster time scale than these nuclear dynamics, we can consider the spins as fixed and thus the value of the nuclear spin state as an exogenous variable. In principle, it is possible to intervene on the nuclear spin state by cooling the system to its ground state and applying tuned frequency pulses set to the transition frequencies of the nuclei. These kinds of pulses will alter the nuclear spin state but not the electron pair spin state (S) or the applied magnetic field (B), and as such can act as an intervention in Woodward's sense. For the bird causal model however, at ambient temperatures, the nuclear spin state is considered an exogenous variable with a fixed value.
- (iii) T : the spatial configuration of the surrounding dominant nuclei. In developing the radical pair model, scientists have considered various molecules as potential candidates for the donor and acceptor molecule. The geometrical arrangement of the nuclei will determine details of the how the nuclear spins couple to the electron pair. The nuclear features of cryptochrome fulfil the spatial requirements to ensure the correct anisotropic coupling required for the model to hold (see Rodgers and Hore 2009).
- (iv) S : the electron pair spin state (considered in the singlet/triplet basis). The value of this variable (again given as a ray in Hilbert space) essentially allows us to calculate the probability that a measurement will return either a singlet or triplet readout at any random time.
- (v) Ys : the final product yield after contact between the reactant molecules. This variable can be considered a binary variable, taking two different values, one for the singlet yield

and one for the triplet yield.

We can intervene on N , B and S , and this is supported by current scientific literature (see previous references, and references therein).

- (i) For reverse engineered analogues of this model, we have the further option of cooling the system, and applying tuned radiofrequency pulses such that we can control the polarity of individual nuclear spins (and thus change the value of N) without directly disturbing either the electron pair spin state or external magnetic field (Robledo et al. 2011). For the bird model, the value of N (the nuclear spin state) is considered fixed. Whilst in principle B has a very small effect on N , in addition to the larger effect on S , this effect is so small it can be approximated to zero, thus there is no causal arrow between B and N . See below for further discussion.
- (ii) Intervention at the level of the electron spin state S , whilst holding N and B fixed, is achieved by the application of tuned, weak, oscillating radiofrequency pulses. These pulses can be tuned to the frequency of the electron pair spin state transitions, and will not disturb the nuclear spin state.
- (iii) And finally, we can intervene on either the direction or the strength of B (treated as a classical variable) whilst holding N fixed, and observe any change in S and Y_s .

What immediately strikes one as surprising about this explanation is that one of the variables (S) is an entangled quantum state of two electron spins. The fact that forces this variable assignment is the final measurement step; the reactant molecules result in different outcomes depending on whether they encounter the radical pair in a singlet or triplet state. It is the difference in total energy that these states possess that determines the different reaction pathways, so in a sense, the reactants are blind to the spin states of each individual electron. This feature circumvents Woodward's worry that we cannot intervene on a joint state of an entangled system. The causal model is not sensitive to the fact that each electron spin state, if measured individually, is randomly distributed. It is the joint state (triplet or singlet) that acts as the difference-maker in this model.

On the basis of these experiments we can derive a simple version of the structural equations relating the variables in this model:

$$P(Y_s) = f(S) \tag{5.38}$$

$$S = f(N, T, B) \tag{5.39}$$

$$N = n \tag{5.40}$$

$$T = t \tag{5.41}$$

$$B = b \tag{5.42}$$

Where the first equation represents a probabilistic dependence and the second equation a deterministic dependence. Typically, causal modelling involves creating either a completely deterministic or probabilistic structure; in this explanation however, the model appears hybridised.

Does this model display the invariance and stability Woodward typically requires of his models? Clearly, yes. There is invariance in so far as the relationships in the model hold for a range of values of B , N and S . If we increase B to extremely high values of field strength we will eventually start flipping nuclear spins and the independence of B and N will no longer hold, so the invariance in this case is limited to a finite range. The model is stable insofar as we can make various changes to background variables, such as atmospheric pressure and temperature, without invalidating the causal entailment of the model.

These dependencies follow the explanation as it is given in much of the literature and are derived from the Hamiltonian dynamics of the system (Cai and Plenio 2013). There are two crucial features of the dynamics that enable the relationships of the causal model to hold. The first is that although, strictly speaking, the Earth's magnetic field exerts an influence on both the nuclear and electron spins, the field is so weak that we can approximate the effect on the nuclear spins to be zero. The Earth's magnetic field has a different effect on electron spin to nuclear spin, essentially because the magneton for nuclear spin and electron spin is very different (by many orders of magnitude). Thus the value for the nuclear spin state remains

approximately fixed for changes in the value of B . This enables an intervention on B to have an effect on S , but not N , thus respecting Woodward's four requirements for an intervention. This kind of idealisation is commonplace in scientific models and has been the focus of much philosophical discussion (see Strevens 2008). Indeed there are many other idealisations that are assumed for this model. For example, the dipole-dipole coupling between the separated electrons is also approximated to zero as the electrons are considered to be sufficiently far apart (see Tiersch and Briegel 2012 for others).

The second feature of the dynamics is best illustrated by considering the standard way to calculate quantum dynamics: expand the initial state (of both subsystems) in the total eigenstates of energy, and the dynamics appears as a phase factor rotating in time for each of the energy eigenstates. The changes in the joint electron spin state that are due to these system dynamics are subsequently witnessed by the chemical reactions that serve to measure the electron joint state in the singlet/triplet basis.

The final causal relation, between the variables S and Y_s , is inherently a probabilistic relation. By varying the joint spin state of the electron pair, we can only change the likelihood that when a reactant molecule encounters the pair it will find it in a singlet state. It is well known that such a measurement step is irreducibly indeterministic and as Woodward endorses probabilistic causal relations, this step should not pose as a problem for his account.

Consider again the main problem Woodward has for quantum causal relations; that it makes no sense to intervene on only one of a pair of entangled states. This worry comes in part from the fact that Woodward uses a notion called modularity (M) to ground the causal Markov condition (CMC) that is necessary for his account (Hausman and Woodward 1999). The modularity constraint requires that the equations that relate the values of different variables in a causal set V are invariant under a range of changes of values of the variables, and is captured by requirements 1–4 in Section 2 above. The concern is that for EPR pairs, modularity fails because the entangled state is inseparable. As Woodward puts it, "...there is no well defined notion of an intervention on the spin state of one of the separated pairs with respect to the other in EPR type experiments." (Woodward 2007, p70). This view hinges to some extent, as Suarez

puts it, “... on an interpretation of the quantum state as non-separable in some rather strong and mysterious ontological sense” (Suarez 2013). We can see that from the explanation of the avian magneto-compass that whilst the nuclear and joint electron spin states are entangled in the hyperfine coupling energy eigenbasis, we are not precluded from intervening on either one or the other variable, and both the experimental verification of the model and the dynamic equations reflect this modularity. The hyperfine coupling Hamiltonian tells us how the nuclear spins affect the electron pair spin dynamics, but not that we must consider both entities as a single variable for the purposes of intervention.

The metaphysical mysteries of coherence, interference and entanglement in this explanation are captured by the dependencies expressed in the structural equations. From an interventionist perspective, of course, we can still identify causal relations and relate without making any metaphysical commitments as to the nature of this causal connection. Quantum engineers who design technology based on this kind of causal identification similarly eschew metaphysical worries and are still able to create devices that exploit such phenomena as entanglement and interference to achieve new ends.

Conclusion

In this paper we do not directly address the traditional problems associated with providing causal explanation for the statistics observed in Bell inequality violating experiments. Rather, we take a different approach to discussing quantum causal explanation. There are recent scientific explanations that appeal crucially to quantum phenomena. The dynamics of these phenomena can be treated in a quantitative manner and used to form causal models. Structural equations can be given based on Hamiltonian dynamics, and interventions performed at each node in the relevant causal diagram. The interventions can be ideal, in the sense that their effects are modelled by numerical simulation, or actualised by experiment. Quantum biology and quantum technology supply many explanations of this kind that could be further explored using this framework. We have given an example from quantum technology, the explanation for LIGO, to illustrate the concept of coherent quantum control, and to provide a qualitative

example of how quantum effects such as interference can play a difference-making role in a macroscopic system. An example from quantum biology, the avian magneto-compass, is also described as a system that utilises difference-making quantum effects. It is presented as a causal-explanatory model using the directed acyclic graph and structural equations method. From a Woodwardian perspective, such explanations are causal. They are stable, invariant and constitute a “small world”. They form an isolated system that can host the external interventions required to test the causal structure of the model.

Acknowledgments I am very grateful to Phil Dowe for his helpful comments, proof reading of the manuscript (in its various forms) and for many informative discussions regarding causation. I am also very grateful to Gerard Milburn for clear discussions regarding the concept of quantum coherence, additional proof reading of the manuscript and for many informative discussions regarding the newest developments in quantum technology.

Bibliography

- Abbott, B., Abbott, R., Adhikari, R., Ajith, P. 2009. “LIGO: the laser interferometer gravitational-wave observatory” *Reports on Progress in Physics* 104, 220502.
- Bandyopadhyay, J.N., Paterek, T., Kaszlikowski, D. 2012. “Quantum coherence and sensitivity of avian magneto reception.” *Physical Review Letters*, 109(11), 110502.
- Cai, J., Plenio, M.B. 2013. “Chemical compass for avian magneto reception as a quantum coherent device” <http://arxiv.org/abs/1304.4143>.
- Cai, J., Guerreschi, G. G., Briegel, H. J. 2010. “Quantum control and entanglement in a chemical compass.” *Physical Review Letters* 104(22), 220502.
- Campbell, R. J., Bickhard, M. H. 2011. “Physicalism, emergence and downward causation” *Axiomathes* 21(1), 33–56.
- Craver, Carl. 2009. *Explaining the Brain*, Oxford University Press.
- Dowling J. P., Milburn, G. J. 2003. “Quantum technology: the second quantum revolution” *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* 361(1809):1655–1674.
- Fleming, G. R., Scholes, G. D., Cheng, Y.-C. 2011 “Quantum effects in biology.” *Procedia Chemistry* 3(1), 38–57
- Frisch, Mathias. 2012. “No place for causes? Causal skepticism in physics.” *European journal for the Philosophy of Science* 2(93):313-336.

- Hausman, Daniel M, and James Woodward. 1999. "Independence, Invariance and the Causal Markov Condition." *The British Journal for the Philosophy of Science* 50: 521–583.
- Huelga, Susana F, and Martin B Plenio. 2013. "Vibrations, Quanta and Biology." *Contemporary Physics* 54(4):181–207.
- Kim, J. 1999. "Making sense of emergence" *Philosophical Studies* 95: 3-36
- Lambert, N., Chen, Y. N., Cheng, Y.-C., Li, C.-M., Chen, G.-Y., Nori, F. 2012. "Quantum biology." *Nature Physics* 9(1), 10–18.
- Milburn, G. J. 2005. *Foundations of quantum technology* *Journal of Computational and Theoretical Nanoscience* 2(2), 161–179.
- Norton, John 2007. "Causation as folk science" In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* eds. Price, H. and Corry, R. Oxford University Press.
- Pearl, Judea. 2000. *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.
- Robledo, Lucio, Lilian Childress, Hannes Bernien, Bas Hensen, Paul F A Alkemade, and Ronald Hanson. 2011. "High-Fidelity Projective Read-Out of a Solid-State Spin Quantum Register." *Nature* 477(7366):574–578.
- Rodgers, C. T., Hore, P. J. 2009. "Chemical magneto reception in birds: the radical pair mechanism." *Proceedings of the National Academy of Sciences* 106(2), 353–360.
- Russell, Bertrand. 1912. "On the Notion of Cause." *Proceedings of the Aristotelian Society* 13:1–26.
- Salmon, Wesley. 1971. *Statistical explanation and statistical relevance*. Pittsburgh: University of Pittsburgh Press.
- Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*. Cambridge, Mass.: MIT Press.

-
- Strevens, Michael. 2008. "Depth: An Account of Scientific Explanation." Harvard University Press. Harvard.
- Suárez, Mauricio. 2007. "Causal inference in quantum mechanics: A reassessment." In *F. Russo J. Williams (Eds.), Causality and probability in the science* (pp. 65–106). London: London College.
- Suárez, Mauricio 2013. Suarez, M. (2013). "Interventions and causality in quantum mechanics" *Erkenntnis* 1–15.
- Tiersch, M., Briegel, H. J. 2012 "Decoherence in the chemical compass: the role of decoherence for avian magneto reception" *Philosophical transactions of the royal society A: mathematical. Physical and Engineering Sciences* 370(1975), 4517–4540.
- van Fraassen, Bas C. 1982. "The Charybdis of Realism: Epistemological Implications of Bell's Inequality." *Synthese* 52:25–38.
- Wood, C. J. and Spekkens, R. W. 2014 "The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning" *New Journal of Physics* 17(3):033002.
- Woodward, James. 2003. *Making Things Happen. A Theory of Causal Explanation*. Oxford: Oxford University Press.
- . 2007. "Causation with a Human Face." In *Causation, Physics, and the Constitution of Reality*, eds. Huw Price, and Richard Corry, 66–105. Oxford: Oxford University Press.
- . 2008. "Mental Causation and Neural Mechanisms" In *Being Reduced: New Essays on Reductive Explanation and Causation*, Eds. Hohwy, J. and Kallestrup, J., Oxford University Press. Chapter 12, 218-262
- Zurek, W. 2014. "Quantum Darwinism, classical reality, and the randomness of quantum jumps", *Physics Today* 67:44.