Quantum Causal Networks

Kathryn Blackmond Laskey

Systems Engineering and Operations Research Department
George Mason University
Fairfax, VA 22030-4444
klaskey@gmu.edu

Abstract

According to the classical nineteenth century worldview, physical systems followed precisely defined trajectories that evolved according to deterministic laws. Physical theory was causally closed, having no place for interventions into its unfolding. Early in the twentieth century, this classical picture was overturned by a new fundamental physical theory. Unlike its classical predecessor, quantum theory is stochastic and causally open. Quantum theory represents not only the passive evolution of closed physical systems, but also the effects of interventions. According to quantum theory, the behavior of a quantum system in response to interventions is intrinsically unpredictable and follows a stochastic law. Stochastic theories of the effects of interventions have become popular recently in artificial intelligence. In these theories, the behavior of an undisturbed system is represented as a graph in which nodes represent variables and directed arcs represent cause and effect relationships. A causal theory specifies both the behavior of the undisturbed system and how it responds to interventions. Interventions act as local surgery to cut the causal links into one or more manipulated variables, and to set the manipulated variables to values specified from outside the model. This paper describes quantum theory as a theory of the effects of interventions, relates it to currently popular theories of causality, and formalizes quantum evolution in terms of graphical probability models defined on density operators.

Introduction

Bohm (1951) said that the quantum state has been called a wave of probability, but it is more accurate to call it a "wave from which many related probabilities can be calculated." In other words, the quantum state predicts not what will occur, nor a single probability distribution for what will occur, but rather a *set* of probability distributions, one for each conceivable intervention that could be made on a quantum system. An intervention results in a stochastic transformation from the state just prior to the intervention to one of the allowable results of the intervention. Quantum theory specifies a probability distribution for the outcome of each such intervention. Thus, quantum theory is naturally viewed as an interventionist theory of causality of the sort that has become popular recently in statistics and the social sciences (Woodward, 2001). Interventionist theories define causal relationships as those in which interventions that change a cause tend to produce changes in the effect. While this view of causality has great intuitive appeal, it has been criticized as being imprecise and potentially circular. One key difficulty is ruling out manipulations that can produce an effect by some means other than the putative cause. A commonly cited example from medicine is the placebo effect, in which administering a drug can produce a cure due to patients' *belief* in the drug's efficacy, regardless of its actual efficacy. It is essential both practically and philosophically to ensure that theories of causality and procedures for drawing causal inferences are not led astray by such spurious effects.

In his seminal book on causality, Pearl (2000) argues that philosophical confusion and the lack of a formal mathematics of causality have hampered our ability to draw sound scientific conclusions about causal relationships. While there is an extensive formal mathematics for the study of logical deduction and statistical association, formal tools for the study of causal relationships have received much less attention. Pearl argues forcefully against the historical tendency among philosophers and scientists to rely on intuition to extract causal conclusions from empirical data. He makes a strong case that formal mathematics is necessary to protect against the biases and errors to which unaided intuition is prone. A formal mathematical theory of causation provides a sound scientific basis for comparing rival causal theories and evaluating their relative degrees of evidential support. Recent developments by Pearl and others (see Woodward, 2001 for references) have gone a long way toward addressing the need for a philosophically coherent and mathematically sound framework for analyzing causal relationships.

Causal claims are stronger than statistical claims. A causal claim asserts not only that the values of two quantities are related to each other, but also that the association is stable under interventions that do not disturb the causal connection. For this reason, Pearl argues that the language of statistical association alone is insufficient for formulating theories of cause and effect relationships, and that new tools and techniques are needed. Drawing on concepts and methods from the theory of structural equations in econometrics and graphical probability models, Pearl has developed a formal mathematics for describing cause and effect relationships, inferring causal relationships from empirical data, predicting the effects of interventions, and drawing inferences about counterfactuals.

Pearl's early writings on causality were based on a probabilistic view of Nature. In his more recent work, Pearl makes an explicit shift toward the Laplacian view of a fundamentally deterministic physical world in which probabilities arise only because of ignorance of boundary conditions:

> …the Laplacian conception is more in tune with intuition. The few esoteric quantum mechanical experiments that conflict with the predictions of the Laplacian conception evoke surprise and disbelief, and they demand that scientists give up deeply entrenched intuitions about locality and causality. Our objective is to preserve, explicate, and satisfy – not destroy – those intuitions.

Pearl argues for a deterministic theory primarily based on its intuitive appeal. Yet, nearly a century of empirical tests have firmly rejected Laplacian determinism in favor of a fundamentally probabilistic alternative. Furthermore, the rival that superseded Laplacian determinism is exactly the kind of theory Pearl says is needed: a formal mathematical theory of the evolution of the behavior of physical systems when subjected to interventions. The purpose of this paper is to explicate quantum theory as a theory of the effects of interventions, to relate it to recent work in the mathematics of causality, and to develop a physically well-founded family of graphical probability models for quantum systems.

The theory presented here differs from Tucci (1995), in that quantum causal networks formalize cause and effect relationships, whereas Tucci's quantum Bayesian networks simply generalize ordinary non-causal Bayesian networks to quantum systems.

Causal Bayesian Networks

In Pearl's theory, a causal model consists of a joint probability distribution over a set of variables, together with a set of "local surgery" rules that specify the effects of intervening to set the states of some of the variables to specified values. The surgery rules amount to cutting the links from causes of a manipulated variable, so that their effects are nullified, and allowing the variable's value to be specified freely via external manipulation. Other causal links, including downstream effects of the variable whose state has been manipulated, are not affected. The result of local surgery is a new joint probability distribution that differs from the original one in that the manipulated variable has a specified state and the rest of the system has a probability distribution determined by the original causal model and the surgery rules that interrupt the normal causal chain to the manipulated variable but leave all other mechanisms undisturbed. Pearl (2000) discusses two kinds of causal model: causal Bayesian networks and structural equation models. In this paper we focus on causal Bayesian netowrks, generalizing them to a class of graphical probability models for quantum systems. Future research will consider quantum analogues for structural equation models.

A *Bayesian network* is a formal representation of the probabilistic relationships among uncertain features of the world. An uncertain feature of the world is represented as a *random variable* X, which takes values in a set $\mathcal{X}$ called the *possible values* for X. A Bayesian network $\mathcal{B} = (G, P)$ represents a joint probability distribution for a collection $X_1, X_2, \ldots, X_n$ of random variables. The first component, G, is a directed graph containing no directed cycles, in which the nodes are in one-to-one correspondence with the random variables. The second component, P, is a collection of local probability models, one for each of the random variables. The random variables that have edges into X_i are called the *parents* of X_i, denoted $\mathrm{pa}(X_i)$. The local probability model for X_i is denoted $\Pr(X_i \mid \mathrm{pa}(X_i))$, and specifies a set of probability distributions for X_i, one for each combination of values of the parents of X_i in G (or a single probability distribution if X_i has no parents). $\Pr(X_i \mid \mathrm{pa}(X_i))$ is specified by defining a rule for obtaining the probability of any possible value $x_i \in \mathcal{X}_i$ as a function of the values of $\mathrm{pa}(X_i)$.

The graph and the local distributions for a Bayesian network define a joint distribution on the random variables as follows:

$$\Pr\left(X_1,\ldots,X_n\right) = \prod_{i=1}^{n} \Pr(X_i \mid \mathrm{pa}(X_i)) \qquad (1)$$

A *causal Bayesian network* (CBN) is a Bayesian network in which the edges represent causal relationships. In Pearl's (2000) formalism, a CBN augments an ordinary Bayesian network with a set of operators $do(X_i{=}x_i)$. The operator $do(X_i{=}x_i)$ is interpreted as a surgical intervention that disconnects X_i from its parents and sets its value to x_i, while leaving the remainder of the causal relationships and local probability models undisturbed. If $V = (X_{i_1}, \ldots, X_{i_k})$ denotes a subsequence of the random variables, and $\Pr^*(X_1, \ldots, X_n \mid do(V{=}\mathrm{v}))$ denotes the probability distribution obtained by applying the $do(\cdot)$ operator to set the random variables in V to the values $v = (x_{i_1}, \ldots, x_{i_k})$, then $\Pr^*(X_1, \ldots, X_n \mid do(V{=}\mathrm{v}))$ is also a Bayesian network. The graph for this new Bayesian network is obtained by deleting from G all edges that point to nodes in V. The local distributions for variables in V place probability 1 on the value set by intervention, i.e., $\Pr^*(X_i = x_i) = 1$ for $i = i_1, \ldots, i_k$. The local probability models for the other random variables are the same as in the undisturbed Bayesian network, i.e., $\Pr^*(X_i \mid \mathrm{pa}(X_i)) = \Pr(X_i \mid \mathrm{pa}(X_i))$ for $i \neq i_1, \ldots, i_k$.

Quantum Theory

Quantum States

States of a quantum system are represented as *density operators* on a Hilbert space associated with the system. A density operator can be identified with a complex-valued square matrix (possibly infinite-dimensional) that is positive and has unit trace, i.e., its eigenvalues are non-negative and sum to one. A density operator is called a *pure* state if it has rank one; otherwise, it is called a *mixed* state. If σ is a rank k density operator on a Hilbert space $\mathcal{H}$, then there exist pure states $\sigma_1, \ldots, \sigma_k$, and positive real numbers $p_1, \ldots, p_k$, such that $\sum_i p_i = 1$ and $\sum_i p_i\sigma_i = \sigma$. For this reason, mixed states can represent uncertainty about the state of a system. That is, $\sum_i p_i\sigma_i$ can represent a system that has probability p_i of being in pure state σ_i. This decomposition into a weighted sum of pure states may not be unique. A state $\sigma = \sum_i p_i\sigma_i = \sum_i r_i\rho_i$ with two different decompositions as probability-weighted sums of pure states could represent either a system having probability p_i of being in state σ_i, or a system having probability r_i of being in state ρ_i. There is no way to distinguish between these possibilities from the state σ alone.

States of composite quantum systems are represented as density operators on tensor product spaces. A tensor product state is the quantum analogue of a Cartesian product state space for classical random variables. A *product state*, written $\sigma_1\otimes\sigma_2\otimes\cdots\otimes\sigma_p$, represents a composite system in which the i^{th} subsystem is in state σ_i. A *separable state* is state that can decomposed as a linear combination $\sum_i p_i\, \sigma_{i1}\otimes\sigma_{i2}\otimes\cdots\otimes\sigma_{ip}$ of product states such the p_i are non-negative real numbers and $\sum_i p_i = 1$. A state of a composite system that is not separable is called *entangled*.

Given a quantum state σ on a tensor product space $\mathcal{H}_1\otimes\cdots\otimes\mathcal{H}_p$, a *reduced density operator* σ_i on the i^{th} Hilbert space can be obtained via an operation called the *partial trace*. More generally, if $i^{(1)}, \ldots, i^{(k)}$ is a subsequence of the integers $1, \ldots, p$, then the partial trace operator can be used to obtain a reduced density operator $S_{i^{(1)}i^{(2)}\ldots i^{(k)}}$ on the Hilbert space $\mathcal{H}_{i^{(1)}}\otimes \mathcal{H}_{i^{(2)}}\otimes\cdots\otimes \mathcal{H}_{i^{(k)}}$. The reduced density operator is the quantum analogue of the marginal distribution for a classical joint distribution. The reduced density operator correctly describes the statistical properties of observable quantities, when attention is restricted to quantities pertaining to the given subsystem.

An important property of entangled systems is that the reduced density matrix can be in a mixed state even when the composite system is in a pure state. When this happens, the subsystem cannot be said to possess a definite state. Mixed states reflecting uncertainty about definite pure states are called *proper mixtures*; mixed states arising from entanglement are called *improper mixtures*. Proper and improper mixtures cannot be empirically distinguished if observations are restricted to those pertaining to the system alone, but can be distinguished if the system and its environment can be observed jointly.

Cerf and Adami (1999) propose a quantum analogue for the classical conditional distribution; Warmuth and Kuzmin (2006) propose a generalization of the Bayesian probability calculus to density matrices. These authors do not address causality.

Unitary Evolution and Stochastic Reduction

Quantum theory as formalized by von Neumann (1955) specifies two kinds of transformations quantum systems can undergo. Passive evolution of an isolated quantum system follows a continuous and reversible process called *Shrödinger* evolution. Given an initial state $\sigma(t_0)$, the state at time $t_1 > t_0$ will be:

$$\sigma(t_1) = U(t_1\text{-}t_0)\ \sigma(t_0)\ \mathrm{U}(t_1\text{-}t_0)^*, \tag{2}$$

where $U(t)$ is a unitary operator given by:

$$U(t) = \exp\{\ \text{-}iHt/\hbar\,\}; \tag{3}$$

H is a Hermitian (i.e., self-adjoint) operator on $\mathcal{H}$ called the *Hamiltonian*; and $\hbar$ is Planck's constant divided by 2π.

The other kind of transformation is a stochastic state change that has been called state reduction, projective measurement, or more picturesquely, collapse. In this paper, the term reduction is preferred because it is more neutral than collapse and applies to a broader class of problems than laboratory measurements. Reduction is represented mathematically as a discontinuous transformation at time t from the state $\sigma(t\text{-})$ to the state $\sigma(t\text{+})$. With a reduction is associated a set $\{P_i\}$ of mutually orthogonal projection operators on $\mathcal{H}$ that sum to the identity, i.e.:

i. $P_i^2 = P_i$;
ii. $P_iP_j = 0$ for $i \neq j$; and
iii. $\sum_i P_i = I$.

The possible outcomes of the reduction are density operators $P_i\sigma(t\text{-})P_i/\mathrm{Tr}(\sigma(t\text{-}))$, where $\mathrm{Tr}(\cdot)$ denotes the trace operator, or sum of diagonal elements of the matrix. Division by $\mathrm{Tr}(\sigma(t\text{-}))$, or normalization, preserves the unit trace property of density operators. Conditional on the time t at which the reduction occurs and the set $\{P_i\}$ of projection operators, the outcome probabilities are given by the Born rule:

$$\begin{aligned}\Pr\big(\sigma(t\text{+}) = P_i\sigma(t\text{-})P_i)/\mathrm{Tr}(\sigma(t\text{-}) \mid \sigma(t\text{-}); \{P_i\}\big)\\ = \mathrm{Tr}(P_i\sigma(t\text{-})P_i)/\mathrm{Tr}(\sigma(t\text{-})).\end{aligned} \tag{4}$$

Because there are at most n mutually orthogonal projection operators of dimension n, the number of possible outcomes of any reduction can be no more than the dimension of the system's Hilbert space. Thus, a density operator on an n dimensional Hilbert space is the quantum analogue of a probability distribution for a random variable with n possible outcomes. Whereas a classical random variable represents outcome probabilities for a single experiment with a given set of n possible outcomes, a density operator represents outcome probabilities for an infinite collection of experiments, each with a different set of n possible outcomes. Quantum probabilities are contingent: *if* the experiment associated with the set $\{P_i\}$ is carried out, *then* the outcome probabilities are given by Equation (4).

Quantum probabilities satisfy a *noncontextuality* property. When a projection operator P has rank less than one minus the dimension of the Hilbert space, there are uncountably many sets of projectors that contain P and satisfy Conditions *i-iii* above. Equation (4) implies that the probability $\mathrm{Tr}(P\sigma P)/\mathrm{Tr}(\sigma)$ of the outcome $P\sigma P/\mathrm{Tr}(\sigma)$ depends only on the projector P and the pre-reduction state σ, and not on the other projectors in the orthogonal set.

Quantum theory as thus formulated is an explicitly temporal theory. Unitary evolution proceeds from past to future. Reductions are instantaneous discontinuous state changes that affect the future evolution of the system but not its past. In relativistic physics, the temporal ordering of two events may depend on the frame of reference. Quantum theory as described in this section is consistent with relativity theory if it is assumed that reductions occur along spacelike surfaces in spacetime (cf. Stapp, 2001).

Quantum theory provides precise predictions for the evolution of isolated systems undergoing Schrödinger evolution and for the probabilities of the outcomes of reductions, but there is no accepted theory for when and how reductions occur. For this reason, intense effort has been devoted to dispensing with reductions by explaining them in terms of unitary evolution of entangled systems. Despite considerable research effort, there remains strong disagreement among physicists about whether this is possible. Because there is no question that von Neumann theory is in accord with observation, and because it provides a natural quantum analogue to classical causal Bayesian networks, we adopt the terminology of reduction in this paper. A deeper debate on the ontological status of reductions is beyond the scope of this paper.

Quantum Operations

Recently, unitary transformations and stochastic reductions have been subsumed into the formalism of *quantum operations*. Quantum operations provide a powerful mathematical tool for representing general transformations of both isolated systems and quantum systems that interact with their environments. The formalism of quantum operations is equivalent to the von Neumann formalism described above, in that any quantum operation can be represented as a composition of unitary operators, stochastic projections and partial traces (Nielsen and Chuang, 2000). Because of their generality and their discrete-time formulation, quantum operations are seeing wide application to analyzing the behavior of quantum systems, especially in quantum computing and quantum information theory.

Quantum operations are especially useful for a theory of quantum causality, because they can describe quantum transformations in which the input and output systems are different. That is, quantum operations can represent interactions in which the behavior of one system has a causal impact on the state of a second system, without requiring an explicit representation of the prior state of the affected system or the post-interaction state of the system producing the effect.

A quantum operation $\mathcal{A}(\sigma)$ is a linear map that transforms operators on an *input* Hilbert space to operators on an *output* Hilbert space, such that the following conditions are satisfied:

1. $\mathrm{Tr}(\mathcal{A}(\sigma)) \leq \mathrm{Tr}(\sigma)$;
2. $\mathcal{A}(\cdot)$ is a *completely positive* map. That is, if σ is a positive operator on the input space, then $\mathcal{A}(\sigma)$ is a positive operator on the output space. Furthermore, if n is a positive integer, ρ is a positive operator on the tensor product of an auxillary n-dimensional Hilbert space and the input space, and $\mathcal{I}_p$ is the identity operator on the auxiliary space, then $(\mathcal{I}_p \otimes \mathcal{A})(\rho)$ is a positive operator.

The partial trace operation that maps a density operator for a composite system to the reduced density operator for a subsystem is an example of a quantum operation. Unitary transformations are also quantum operations. If P is a projection operator, the map from σ to $P\sigma P$ is a quantum operation that does not preserve the trace. Trace-preserving quantum operations correspond to deterministic state transitions or stochastic transitions in which the outcomes are not distinguishable. Trace-reducing quantum operations represent stochastic transformations with distinguishable outcomes. Consider a set $\mathcal{A}_1(\cdot), \ldots, \mathcal{A}_n(\cdot)$ of trace-reducing quantum operations such that $\sum_i \mathrm{Tr}(\mathcal{A}_i(\sigma)) = \mathrm{Tr}(\sigma)$ for all σ. This set represents a process in which a transformation is chosen by a stochastic rule. The probability that the i^{th} transformation occurs is given by $\mathrm{Tr}(\mathcal{A}_i(\sigma))$, and the result of the i^{th} transformation on input σ is $\mathcal{A}_i(\sigma)/\mathrm{Tr}(\sigma)$. In particular, state vector reduction with orthogonal projector set $\{P_i\}$ is an example of a quantum operation with a stochastic outcome.

To bring the theory of quantum operations into concordance with relativity theory, the output system for a quantum operation must be localized in a region of spacetime that does not overlap the past light cone of the input system. For stochastic projection operations, the output system may have a spacelike separation from the input system. For time evolution quantum operations, the output system must be localized within the future light cone of the input system.

Fiducial Projections

When the state space has dimension n, there exists a set $F_1, \ldots, F_{n^2}$ of projection operators, such that the state is characterized by the Born probabilities associated with the F_i (Nielson and Chuang, 2000; Hardy, 2002). Any such collection $\{F_i\}$ is called a set of *fiducial* projections (Hardy, 2002). If $\{F_i\}$ is a fiducial set, and σ and ρ are two density operators such that $\mathrm{Tr}(F_i\sigma F_i) = \mathrm{Tr}(F_i\rho F_i)$ for $i = 1, \ldots, n^2$, then $\sigma = \rho$. The fiducial projections can be chosen to have rank 1. In this case, the fiducial projections are themselves density operators, and they represent pure states of the system. Because F_i is a projection operator with rank 1, it can be shown that if $F_i\sigma F_i \neq 0$, then $F_i\sigma F_i/\mathrm{Tr}(F_i\sigma F_i) = F_i$.

A fiducial projection operator F_i thus represents both a pure state of the system and an intervention that has F_i as one of its possible outcomes. If the intervention F_i is applied to a system whose pre-intervention state is σ, then the probability is $\mathrm{Tr}(F_i\sigma F_i)$ that the post-intervention state is to F_i. Because of noncontextuality, these probabilities hold for *any* intervention in which F_i is one of the possible outcomes, regardless of the other possible outcomes of the intervention.

Just as quantum states can be characterized by the probabilities associated with fiducial operators, quantum operations can be characterized by how they act on fiducial operators. Specifically, let $F_1, \ldots, F_{n^2}$ be a set of fiducial projectors on an n-dimensional input Hilbert space and let $G_1, \ldots, G_{m^2}$ be a set of fiducial projectors on an m-dimensional output space. Suppose that $\mathcal{A}(\cdot)$ and $\mathcal{A}'(\cdot)$ are completely positive maps such that $\mathrm{Tr}(G_j\mathcal{A}(F_i)G_j) = \mathrm{Tr}(G_j\mathcal{A}'(F_i)G_j)$ for $i=1,\ldots,n$ and $j=1,\ldots,m$. Then $\mathcal{A}(\cdot)$ is equal to $\mathcal{A}'(\cdot)$ (Nielsen and Chuang, 2000, sec. 8.4.2).

Graphical Models for Quantum Systems

Sequenced Association Graphs

Sequenced association graphs are proposed as a quantum analogue to the acyclic directed graphs used to model dependence relationships in CBNs. Sequenced association graphs represent allowable kinds of dependencies for quantum systems.

In a CBN, the arcs are directed and the probabilistic dependencies are causal. Of course, it is easy to find real-world examples of correlations that do not correspond to causal relationships. Nevertheless, outside the quantum realm, it is generally assumed that Riechenbach's principle of common causes holds. That is, when two quantities are correlated, it is assumed either that one is a cause of the other or that there is another variable that is a common cause of both. When the principle of common cause holds, one can construct a CBN by inserting hidden variables to represent common causes of correlated variables.

In quantum systems, although entanglement can give rise to correlations between spacelike separated events, causal influence can operate only between timelike separated events, and only from past to future. This fundamental difference between correlations involving spacelike and timelike separated events is represented in sequenced association graphs by using directed arcs to represent causal influences from the past to the future, and undirected arcs to represent correlations between contemporaneous entangled systems.

***Definition* 1:** Let G be a graph, and let A and B be nodes of A. Then A and B are *contemporaneous* if (i) there is an undirected edge connecting A and B, or (ii) there is an undirected edge between A and a node contemporaneous with B. If A and B are contemporaneous, we write $A \sim_{\mathrm{T}} B$.

***Definition* 2:** Let G be a graph, and let A and B be nodes of A. Then *A precedes B* if (i) there is a directed edge from A to B, or (ii) there is a directed edge from A to a node that precedes B. If A precedes B, we write $A \prec_{\mathrm{T}} B$.

A straightforward inductive argument shows that $\sim_{\mathrm{T}}$ is an equivalence relation and $\prec_{\mathrm{T}}$ is transitive.

***Definition* 3:** A graph G is a *sequenced association graph* (SAG) if there is no pair of nodes A and B such that (i) A precedes B and (ii) B precedes or is contemporaneous with A.

The directed arcs in a sequenced association graph establish a partial order on the nodes. When a SAG is used to model a physical process, each node is associated with a physical system localized within a region of spacetime. Directed edges connect timelike separated systems, and are oriented from past to future. Undirected edges connect spacelike separated systems that are correlated due to entanglement.

Because contemporaneity is an equivalence relation, it partitions the nodes of a SAG into equivalence sets. The elements of this partition are called CN-sets.

***Definition* 4:** Let G be a sequenced association graph. A *CN-set* is a maximal subset of mutually contemporaneous nodes of G. A *root CN-set* is a CN-set in which none of the arcs in G enters any of the nodes in the CN-set. A CN-set that is not a root CN-set is called a *child CN-set*.

Figure 1 shows a SAG containing five CN-sets, enclosed in dotted ovals and numbered 1A through 4. The numbering scheme indicates the time order if it can be established from the graph. Letters are appended to the numbers to label nodes for which the order cannot be distinguished. The time ordering of CN-sets 1A and 1B cannot be determined from the graph; the CN-sets 2 through 4 follow these sets in temporal order.

A Simple Two-Node Quantum Causal Network

A quantum causal network (QCN) is proposed as a quantum analogue to a CBN. Like a Bayesian network, a QCN uses a graph to represent qualitative relationships and local probability models to represent numerical likelihood information. Whereas the graph for a CBN is an acyclic directed graph, the graph for a QCN is a sequenced association graph. Density matrices and quantum operations represent numerical likelihood information in a QCN.

To introduce the fundamental concepts, we consider a simple two-node graph $X \to Y$. The state space for a CBN with this graph is defined by specifying a set of possible values for each of the nodes X and Y. For a QCN, Hilbert spaces $\mathcal{H}_X$ and $\mathcal{H}_Y$ are specified for the nodes X and Y. These Hilbert spaces have dimension equal to the maximum cardinality of the outcome set of an intervention that could be performed at the corresponding node.

The joint distribution for the undisturbed CBN is specified by defining a probability distribution $\Pr(X)$ for X, and a set $\Pr(Y|X)$ of probability distributions for Y, one for each possible value of X. It is assumed that exactly one pair of possible values will occur. The probability of the value $(X, Y) = (x, y)$ is given by $\Pr(x)\Pr(y|x)$. The density operator for the undisturbed QCN is specified by defining density operator σ_X on $\mathcal{H}_X$ and a trace-preserving quantum operation $\mathcal{A}_{Y|X}(\cdot)$ that maps $\mathcal{H}_X$ to $\mathcal{H}_Y$. This defines a density operator on the product space $\mathcal{H}_X \otimes \mathcal{H}_Y$ as follows. It is well known that any density operator can be expanded as a mixture of mutually orthogonal 1-dimensional density operators (cf., Nielsen and Chuang, 2000). Thus, we can write

$$\sigma_X = \sum_i \theta_i Q_i, \qquad (5)$$

where the Q_i are mutually orthogonal one-dimensional projection operators on $\mathcal{H}_X$, and the θ_i are non-negative numbers that sum to 1. We can also write

$$\mathcal{A}_{Y|X}(Q_i) = \sum_j \rho_{ij} R_{ij}, \qquad (6)$$

where for each i, the R_{ij} are mutually orthogonal one-dimensional projection operators $\mathcal{H}_Y$, and the ρ_{ij} are non-negative numbers that sum to 1. Note that the mixture components R_{i1}, R_{i2}, … for $\mathcal{A}_{Y|X}(Q_i)$ may be different for different i. Now, we can form a joint density operator on the tensor product space as follows:

$$\tau_{XY} = \sum_{i,j} \theta_i \rho_{ij} Q_i \otimes R_{ij}. \qquad (7)$$

The density operator τ_{XY} represents a quantum state for the undisturbed two-node QCN. Applying the partial trace yields density operators σ_X and σ_Y representing the states of the X and Y subsystems of the undisturbed joint system.

An intervention to change X in a CBN would apply $do(X{=}x)$ to replace $\Pr(X)$ with a new joint distribution $\Pr^*(X)$ in which X has value x with certainty. The marginal distribution of Y would become $\Pr(Y|X{=}x)$. In contrast, intervening to change X in a QCN means initiating a reduction. That is, we specify a set $P_1, \ldots, P_n$ of orthonormal projectors (i.e., operators satisfying *i-iii* above). The outcome of the intervention is stochastic, with possible outcomes $\sigma_i = P_i\sigma_X P_i/\mathrm{Tr}(P_i\sigma_X P_i)$ $i = 1, \ldots, n$. The outcome probabilities are given by the Born rule (4). If the outcome is $P_i\sigma_X P_i/\mathrm{Tr}(P_i\sigma_X P_i)$, then the original CBN is replaced by a new CBN in which the quantum operation $\mathcal{A}_{Y|X}(\cdot)$ associated with Y remains unchanged, and the density operator associated with X becomes $P_i\sigma_X P_i/\mathrm{Tr}(P_i\sigma_X P_i)$.

The probability distribution for an n-state root node of a Bayesian network can be defined by specifying n-1 real numbers. The density matrix for an n-dimensional root node of a QCN can be defined by specifying n^2-1 real numbers. In both cases, one degree of freedom is subtracted to account for the constraint that probabilities sum to 1. The conditional distribution for an m-state child node with an n-state parent can be defined by specifying $n(m$-1) real numbers. A trace-preserving quantum operation from an n-dimensional Hilbert space to an m-dimensional Hilbert space can be defined by specifying $n^2(m^2$-1) real numbers. As above, degrees of freedom are subtracted to account for the normalization constraint.

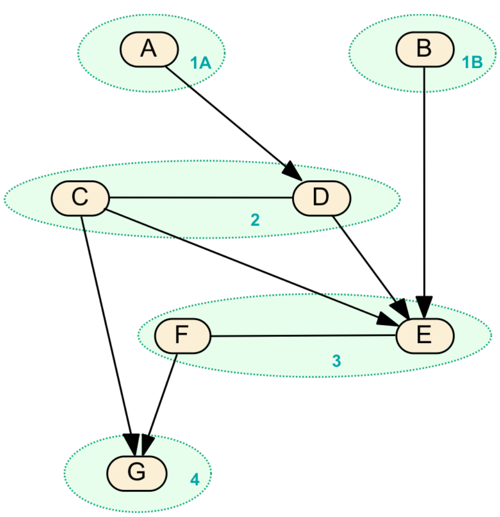

Figure 1: Sequenced Association Graph

General Quantum Causal Networks

A general theory of quantum causal networks extends the two-node example of the previous section to sequenced association graphs having arbitrary number of nodes.

***Definition* 5:** Let G be a SAG, and let $\{X_1, \ldots, X_k\}$ be a child CN-set for G. A node Y is an *influencing parent* for the CN-set if G has a directed edge from Y to one of the X_i, and a *non-influencing parent* for the CN-set if it is contemporaneous to a parent for the CN-set.

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

1. A density operator $\Delta(X_1, \ldots, X_k)$ on $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$ if $\{X_1, \ldots, X_k\}$ is a root CN-set;
2. A quantum operation $\Delta(X_1, \ldots, X_k \mid W_1, \ldots, W_r)$ from $\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_r$ to $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$ if $\{X_1, \ldots, X_k\}$ is a child CN-set.

***Definition* 7:** Let $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ be a product space. A *fiducial reduction* is a set of projection operators satisfying conditions *i-iii*, in which each projector in the set is a product $F_1 \otimes \cdots \otimes F_n$ of fiducial projectors.

***Definition* 8:** Let G be a SAG, and let $\{X_1, \ldots, X_k\}$ be a root CN-set. The local distribution $\Delta(X_1, \ldots, X_k)$ *respects G* if for any fiducial reduction applied to $\Delta(X_1, \ldots, X_k)$ and any i, the conditional probability of X_i given $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$ depends only on the neighbors of X_i in G.

***Definition* 9:** Let G be a SAG, let $\{X_1, \ldots, X_k\}$ be a child CN-set, and let $\{W_1, \ldots, W_r\}$ denote its influencing and non-influencing parents. The local distribution $\Delta(X_1, \ldots, X_k \mid W_1, \ldots, W_r)$ *respects G* if the following condition holds. For i=1,…, r, let F_i denote a fiducial projector on the Hilbert space for W_i. Let $\Delta(X_1, \ldots, X_k \mid W_1, \ldots, W_r)(F_1 \otimes \cdots \otimes F_r)$ denote the quantum operation $\Delta(X_1, \ldots, X_k \mid W_1, \ldots, W_r)$ applied to the product projector $F_1 \otimes \cdots \otimes F_r$. Then the conditional probability assigned by $\Delta(X_1, \ldots, X_k \mid W_1, \ldots,$

$W_r)(F_1\otimes\cdots\otimes F_r)$ to X_i given $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$ depends only on those X_j that are neighbors of X_i in G and those W_j that are parents of X_i in G.

***Definition* 10:** Let G be a sequenced association graph. Let $\{\mathcal{H}_i\}$ be a collection of Hilbert spaces, one for each node X_i of G. Let $\{\Pr(\cdot)\}$ be a set of local distributions, one for each CN-set of G. Then $\mathcal{Q} = (G, \{\mathcal{H}_i\}, \{\Pr(\cdot)\})$ is a *quantum causal network* if each of the local distributions respects G.

The density operator for a root CN-set of a QCN requires at most n^2-1 real numbers to specify, where n is the product of the dimensions of the Hilbert spaces for the nodes in the CN-set. The quantum operation for a child CN-set requires at most $n^2(m^2$-1) real numbers, where m is the product of the dimensions of the Hilbert spaces for the parent nodes and n is the product of the dimension of the nodes in the child CN-set. The independence assumptions encoded in G reduce the number of parameters needed to specify these local distributions.

As for the two-node example described above, a general QCN induces a density operator at each of its nodes. Propagating the quantum operations forward in the direction of the causal arcs induces a density operator on each CN-set. A generalization of the construction (5) – (7) can be applied to construct a density operator on the tensor product space. A reduced density matrix for each node can be obtained via the partial trace operation. These density matrices represent undisturbed evolution of the quantum system. Undisturbed evolution changes the state deterministically, although non-unitary quantum operations in which the input and output state spaces are the same increase quantum entropy (Nielsen and Chuang, 2000).

Interventions General QCNs

If $\mathcal{Q}$ is a QCN, an intervention at a single target node T is modeled as follows. Let CN(T) denote the CN-set of T. Let $\mathcal{H}_{\mathrm{CN}(T)}$ denote the associated Hilbert space, and let $\sigma_{\mathrm{CN}(T)}$ denote the reduced density operator for CN(T). Let $P_1, \ldots, P_n$ be a set of orthonormal projection operators (i.e., satisfying Conditions *i-iii* above) on $\mathcal{H}_{\mathrm{CN}(T)}$, such that each P_i acts as the identity on all nodes except T. The intervention results in a new QCN $\mathcal{Q}^*$, where:

1. The graph G^* of $\mathcal{Q}^*$ is obtained from graph G of $\mathcal{Q}$ by removing all directed arcs entering CN(T).
2. The new local distribution for CN(T) is chosen stochastically. The possible values are $\sigma_i = P_i\sigma_{\mathrm{CN}(T)}P_i/\mathrm{Tr}(P_i\sigma_{\mathrm{CN}(T)}P_i)$, for $i = 1, \ldots, n$. The probability of obtaining σ_i is $\mathrm{Tr}(P_i\sigma_{\mathrm{CN}(T)}P_i)$. Note that all independence relationships among nodes in CN(T) that existed in $\sigma_{\mathrm{CN}(T)}$ are preserved in σ_i. Therefore, σ_i respects G^*.
3. The local distributions for all nodes not in CN(T) are unchanged.

As a result of the intervention, the target node takes on one of the allowable results for the projection set associated with the reduction operation. If the target node is entangled with contemporaneous neighbors, intervening at the node may affect these neighbors even though the projector acts as the identity on these nodes. Post-intervention states for descendants of the target node's CN-set are obtained by forward propagation.

To see how this works, consider the example of a pair of qbits in an entangled state. Let $\uparrow$ (up) and $\downarrow$ (down) denote two orthogonal states for the qbits. Suppose the system begins in a pure state having equal amplitude on the $\uparrow\downarrow$ and $\downarrow\uparrow$ states. The reduced density matrix for each individual qbit is a mixed state with 50% weight on $\uparrow$ and 50% weight on $\downarrow$. This means that an intervention to force the first qbit into either the $\uparrow$ or the $\downarrow$ state will give a 50% chance for each of the two possibilities. After reduction, the second qbit is certain to have value opposite to the first. That is, the intervention result will be $\uparrow\downarrow$ with 50% probability and $\downarrow\uparrow$ with 50% probability. Both possible outcomes of the intervention are different from the original entangled state. Nevertheless, if we do not condition on which outcome has occurred, the conditional probability that a projection of either qbit onto $\uparrow$ or $\downarrow$ will yield the $\uparrow$ state remains at its pre-reduction value of 50%. Physically, the original situation represents a pure state of an entangled system, in which reduced density operators for the individual qbits are mixed states that place 50% weight on each alternative. These reduced density operators do not correspond to true mixtures. Rather, the 50% weights represent conditional probabilities that *if* an intervention forces the qbit into the $\uparrow$ or $\downarrow$ state, each possibility will occur with 50% probability. The post-reduction state represents a true mixture. The reduction has already forced a choice between the $\uparrow$ or $\downarrow$ states, but which of these has occurred has not been specified. The global 2-qbit system is no longer in a pure entangled state, but is in an unknown product state. These two possibilities, entangled pair or unknown product state, cannot be distinguished by observing either qbit in isolation. However, the behavior of the composite 2-qbit system is different in the two cases. Entanglement is responsible for many of the most interesting aspects of quantum systems. It is believed that quantum computers are intrinsically more powerful than classical computers, and entanglement is the source of this power.

Control Through Intervention

According to quantum theory, the kind of intervention represented by Pearl's $do(X{=}x)$ operator, in which a random variable is set to a specified value, is not physically realizable. If a random variable X has value x initially, then any quantum intervention in which x is one of the possible outcomes will result in x with probability 1. An intervention changes the state only when there are several possible outcomes that are not orthogonal to the initial state.

In his proof that the entropy of a pure quantum state is zero, von Neumann (1955; chapter V.2) showed how to transform a pure state x to an orthogonal pure state y by applying a sequence of projectors in rapid succession, each of which superposes the states x and y, and in which the magnitude of the weight on y increases as the sequence progresses. Although almost all treatments of interventions in the quantum theory literature are heuristic and informal, the ability to control the behavior 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 analyzing the effects of interventions has sowed confusion and hindered advances in practical applications of quantum theory. As a formal theory of the effects of interventions, QCNs are a useful tool for analyzing quantum systems and their behavior.

The predictions of quantum theory have been subjected to extensive empirical testing for a wide variety of quantum processes, with stunning agreement between theory and empirical results. However, quantum theory as presently formulated contains a major explanatory gap. The theory has nothing at all to say about when a reduction will occur and which set of orthogonal projection operators will correspond to the possible results. Despite intense effort over many years, no one has yet found a satisfactory way to dispense with reductions and still bring quantum theory into concordance with the results of measurements, and physicists disagree strongly about the feasibility of the endeavor. Because reductions are associated with scientists performing measurements, the lack of a theory for state reduction has been called the "measurement problem."

Rather than attempt to dispense with reductions, the approach taken in this paper is to formalize reductions as external interventions in a causal graphical model formulation of quantum theory. The content of the theory described here fully consistent with standard von Neumann / Copenhagen quantum theory, but it is explicated in a language that ties it firmly to recent work on probabilistic models of causality. It is hoped that formulating a quantum version of causal graphical models will shed light on the physical realizability of causal theories.

In particular, Pearl's *do*-calculus can be viewed as a classical approximation to a more physically realistic quantum theory of causation. One role for a quantum theory of causality is to explicate conditions under which such an approximation is adequate. A Pearl-style causal Bayesian network is a reasonable approximation when: (1) decoherence effects nearly eliminate the off-diagonal elements of the density operator for all subsystems under consideration, rendering the global system essentially equivalent for all practical purposes to a statistical mixture of quasi-classical states; and (2) it is possible to apply a von Neumann style sequence of operators in rapid succession to drive the state of any subsystem to any desired state, without major disturbance to other subsystems. The first condition holds in many cases of practical interest, but the second condition may be more problematic.

A more physically realistic quantum theory of causation may open up new avenues of investigation. Specifically, it opens the door to new, theoretically well-founded research into the kinds of interventions that are physically achievable the conditions under which they can be applied without disturbing the states of and causal interactions among subsystems other than the targets of intervention.

Discussion

A quantum state is defined as a set of potentialities, that is, conditional probabilities for the results of any conceivable interventions that can be applied to it. Thus, quantum theory is at its core an interventionist causal theory of the sort recently popularized by Pearl and others. The formalism of quantum causal networks provides a language for representing cause and effect interactions among quantum systems, and for posing questions about the effects of interventions. It is anticipated that the theory will prove useful for analyzing and designing quantum computing devices and algorithms. Artificial intelligence systems must be implemented in hardware, and physical hardware obeys the laws of quantum theory. Quantum graphical probability models may ultimately replace classical computability theory as a theoretical foundation for an artificial intelligence grounded firmly in physical theory.

References

Bohm, D., 1951. *Quantum Theory*. New York: Prentice-Hall.

Cerf, N.J. and Adami, C., 1999. Quantum extension of conditional probability. *Physical Review A*, **60**(2):893–897.

Hardy, L., 2001. *Quantum Theory from Five Reasonable Axioms*. http://arxiv.org/abs/quant-ph/0101012.

Nielsen, M.A. and Chuang, I.L., 2000. *Quantum Computation and Quantum Information*. Cambridge, UK: Cambridge University Press.

Pearl, J., 2000. *Causality*: *Models, Reasoning and Inference*. Cambridge, UK: Cambridge University Press.

Stapp, H., 2001. Quantum Mechanics and the Role of Mind in Nature. *Foundations of Physics 31*, 1465-99.

Tucci, R., 1995. *Quantum Bayesian Nets*. http://arxiv.org /abs/quant-ph/9706039.

Warmuth, M. and Kuzmin, D., 2006. A Bayesian Probability Calculus for Density Matrices. *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, San Mateo, CA: Morgan Kaufmann.

Woodward, J., "Causation and Manipulability", The Stanford Encyclopedia of Philosophy (2001), E. N. Zalta (ed.), <http://plato.stanford.edu/archives/win2003/entries /causation-mani//>

von Neumann, J., 1955. *Mathematical Foundations of Quantum Mechanics*. Princeton, NJ: Princeton University Press.