Quantum Physical Symbol Systems

KATHRYN BLACKMOND LASKEY
Department of Systems Engineering and Operations Research, MS4A6, George Mason University,
Fairfax, VA 22030, USA
E-mail: klaskey@gmu.edu

Abstract. Because intelligent agents employ physically embodied cognitive systems to reason about the world, their cognitive abilities are constrained by the laws of physics. Scientists have used digital computers to develop and validate theories of physically embodied cognition. Computational theories of intelligence have advanced our understanding of the nature of intelligence and have yielded practically useful systems exhibiting some degree of intelligence. However, the view of cognition as algorithms running on digital computers rests on implicit assumptions about the physical world that are incorrect. Recently, the view is emerging of computing systems as goal-directed agents, evolving during problem solving toward improved world models and better task performance. A full realization of this vision requires a new logic for computing that incorporates learning from experience as an intrinsic part of the logic, and that permits full exploitation of the quantum nature of the physical world. This paper proposes a theory of physically embodied cognitive agents founded upon first-order logic, Bayesian decision theory, and quantum physics. An abstract architecture for a physically embodied cognitive agent is presented. The cognitive aspect is represented as a Bayesian decision theoretic agent; the physical aspect is represented as a quantum process; and these aspects are related through von Neumann's principle of psycho-physical parallelism. Alternative metaphysical positions regarding the meaning of quantum probabilities and the role of efficacious choices by agents are discussed in relation to the abstract agent architecture. The concepts are illustrated with an extended example from the domain of science fiction.

Key words: Bayesian networks, decision theory, graphical models, interpretations of probability, quantum computing, quantum measurement

1. Introduction

The Information Age has brought about fundamental changes in the way humanity approaches the study of mental and cognitive phenomena. There has been a move toward the development of theories of mind that can be implemented as computer programs, and yield predictions that can be compared with empirical data (cf., Newell and Simon, 1976; Anderson, 1999; Thagard, 1988). Newell and Simon (1976) argued that intelligence is a property possessed by physical agents operating in a physical environment, whose cognition is performed by a physical brain and nervous system. Therefore, the means by which an agent identifies, evaluates, selects among, and implements plans is constrained by the laws of physics. They pioneered the methodological approach of implementing theories of

intelligence in physical computing devices that can carry out the required computations and actions rapidly and accurately enough to allow the agent to operate successfully in its environment. Newell and Simon viewed intelligence as a property possessed to greater or lesser degree by a *physical symbol system*, which they defined as:

> ...a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). [Symbols in a structure] are related in some physical way... A physical symbol system ... produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider than just these symbolic expressions themselves.

Two essential capabilities a physical symbol system must possess are *designation* and *interpretation*. Symbol structures can designate objects in the world external to the system, thus allowing the system to affect and/or be affected by the designated object. Symbol structures designating a sequence of actions can be interpreted, thus allowing the system either to act in the world or to issue instructions to control the actions of an external system. Given this definition, Newell and Simon articulated the *physical symbol system hypothesis*:

> A physical symbol system has the necessary and sufficient means for intelligent action.

They offered the physical symbol system hypothesis as a scientific hypothesis about the nature of intelligence. They challenged the community to develop the hypothesis further and subject it to empirical test. The intervening years have seen a vigorous debate on the physical symbol system hypothesis. Attempts to construct physical devices that behave intelligently have resulted in partial success, in vastly improved understanding of the nature of the challenge, and in a better appreciation of what we do and do not understand about intelligence and its role in Nature. The resulting cross-fertilization between artificial intelligence, biology and cognitive psychology has led to advances in all these disciplines.

As cognitive scientists have worked to develop physically grounded theories of cognition, physicists have discovered the necessity of bringing cognitive agents into physical theory. Prior to the twentieth century, science focused on constructing and evaluating theories of a physical world evolving independently of the theorizing agent. The relationship of the domain under investigation to the mind that formulates and tests scientific theories was regarded as outside the province of science. This separation has been rendered untenable by the advent of quantum theory, the most stunningly successful scientific theory humanity has yet achieved. In a radical departure from classical physics, an observer in quantum theory enters into the dynamics of a physical system in a fundamental way. According to the

orthodox interpretation of the theory, known as the Copenhagen interpretation, the observable behavior of a quantum system depends on whether an agent decides to observe it, and if so, on when it is observed and what features the agent chooses to observe. Quantum theory also departs from the strict determinism of classical physics. Although statistics is important in classical statistical mechanics, it is regarded as an approximation due to imperfect knowledge of the details of systems whose underlying dynamics are deterministic. Uncertainty is a fundamental aspect of quantum theory: the observable behavior of physical systems follows intrinsically probabilistic laws. In addition, the act of acquiring knowledge about a system affects the system under observation in ways that can be predicted only imperfectly.

The developers of the orthodox Copenhagen interpretation viewed quantum theory as a theory of the knowledge of observers. They explicitly eschewed an ontology referring to a physical reality external to the mind of the observer. According to Heisenberg (Heisenberg, 1958), "The conception of objective reality of the elementary particles has thus evaporated. . . into the transparent clarity of a mathematics that represents no longer the behavior of particles but rather our knowledge of this behavior." The view that quantum theory is about the knowledge of observers and not about an external physical reality has been controversial since its inception. There have been numerous attempts to reformulate the theory to remove observer dependence. While some of these attempts have gained a substantial following, and passionate debate over the foundations of quantum theory continues, the Copenhagen interpretation remains standard.

To summarize, an influential school of thought within the computer and cognitive sciences views physical embodiment as a fundamental property of intelligence. At the same time, the orthodox interpretation of quantum physics gives the observer an essential role in physical theory. This state of affairs suggests that a unified science of cognition and information processing must encompass not only the means by which agents represent and reason about the world, but also the physical processes by which agents create and revise their representations, as well as the manner in which the physical and cognitive aspects of intelligence relate to each other. Intelligent agents form, manipulate and evolve representations of the world they live in. If they are to survive and flourish, their representations must yield sufficiently accurate predictions to enable them to identify and pursue life-enhancing courses of action. For this to be possible, the physical world must be reasonably predictable, and must also permit the evolution of physical systems capable of forming representations. Thus, theories of knowledge representation, learning and action selection must respect the constraints physics imposes on the interface between the physical world and the agents who acquire information about and act upon the world.

This paper argues that a marriage of quantum theory with Bayesian decision theory provides a unifying account of the physical and informational aspects of physical symbol systems. Section 2 provides an introduction to graphical probability and decision models. Graphical models have become increasingly popular as a language for expressing logically sound and tractable computational domain

theories, for representing cause and effect relationships, and for specifying inference and optimization algorithms. Section 3 describes an interpretation for quantum theory that maps directly to Bayesian decision theory. A representation of quantum evolution and quantum measurement is provided in the language of graphical models. With no changes to the mathematical structure of existing physics, the ontology proposed here connects physical reality in a plausible way to efficacious choices by agents. Section 4 brings these threads together in a theory of quantum cognitive agents – that is, physically embodied agents in which cognition is modeled as a physical quantum process related to a sequential Bayesian decision process via the principle of psycho-physical parallelism.

2. Graphical Probability and Decision Models

Intelligence requires the ability to reason and act in the presence of uncertainty. One of the most difficult challenges in artificial intelligence has been the development of principled yet tractable methods for plausible reasoning and decision making in the presence of uncertainty and incomplete information. Although once controversial, Bayesian decision theory is now regarded as a foundational theory for computational inference and decision under uncertainty, and has become a standard of comparision for proposed alternative approaches (c.f., Russell and Norvig, 2002). Game theory, or multi-agent decision theory, is becoming standard as a foundation for systems of multiple interacting agents (e.g., Kearns and Mansour, 2002).

2.1. Bayesian Decision Theory

Bayesian decision theory is a mathematical theory of rational decision making under uncertainty. The theory provides a sound way to combine beliefs with values to arrive at logically consistent, value-driven decisions. It applies to situations in which an agent must choose an action or series of actions from among a set of alternatives. The consequence of the choice depends on both the selected action(s) and the state of the world. The state of the world, which may be unknown to the agent at the time the choice is made, belongs to a set of mutually exclusive and collectively exhaustive possible states. A decision theoretic agent expresses uncertainty about the state of the world as a probability distribution, and expresses preferences among consequences as a utility function. Taken together, the possible actions, possible states, possible consequences, and the probability and utility functions comprise the agent's decision theoretic model. According to the model, the agent's optimal choice is to select the action for which the mathematical expectation of the utility is maximized. When the agent acquires information about the world, probability assignments are updated according to a mathematical formula called *Bayes rule*. The revised probabilities are used for subsequent predictions and decisions.

Although many textbooks take states, consequences, acts, probabilities and utilities as givens, it is possible to construct a decision theoretic model from more cognitively natural primitives. A number of authors (e.g., Savage, 1954; Pratt et al., 1965) have developed axiom systems that capture intuitive notions of rational behavior, and demonstrated that the axioms imply the optimality of subjective utility maximization. For example, in Savage's (1954) system, the primitives are states, consequences, acts (represented as functions from states to consequences), and preferences among acts. Agents whose preferences satisfy a set of rationality axioms make choices "as if" they are maximizing the expected value of a subjective utility function over preferences, where the expectation is taken with respect to a subjective probability distribution over choices. Thus, probabilities and utilities are derived from preferences over acts. A common criticism of Savage's system is the requirement of a global preference ordering over acts. Savage himself expressed concern that although his theory postulates a global preference ordering, agents typically restrict consideration to isolated, simplified microcosms, neglecting features of the world not related to the problem at hand. He called this the problem of "small worlds." Recent work on computational decision theory attempts to build formal decision models in which microcosm preferences are taken as primitive and the "grand world" decision model is constructed by composing "small world" models (c.f., Shafer, 1981; Blume et al., 2005).

Many arguments have been put forward both for and against the principle of maximum expected utility as a model of rational decision making under uncertainty (e.g., Howson and Urbach, 1993). A common objection to axiomatic arguments is that computing optimal policies for realistically complex decision problems is intractable.[1] Great strides have been made in recent years (e.g., Pearl, 1988; Jensen, 2001; Neapolitan, 2003; Korb and Nicholson, 2003) in tractable exact or approximate algorithms for computational probabilistic inference and decision making. Decision theoretic methods have found their way into numerous successful applications (e.g., Heckerman et al., 1995; Levitt et al., 1995; Parker and Miller, 1987). Among pragmatists, successful applications provide a stronger argument in favor of Bayesian decision theory than axiomatic arguments. It has been argued that intelligence requires the functional equivalent of approximate Bayesian inference and decision theory (e.g., Lee and Mumford, 2003). Many heuristic methods proposed as alternatives to decision theory can be shown to result in approximate decision theoretically optimal behavior within their domain of applicability (e.g., Martignon and Laskey, 1999). When computational limits are taken into account, decision theory itself would recommend an approximately optimal heuristic strategy over an

---

[1] Other counter-arguments question some of the standard axioms, such as those implying a simple ordering of all options. Relaxing the objectionable axioms has led to alternative decision theories, such as theories of interval probabilities and utilities. These theories are not discussed here, except to note that they typically pose even more challenging computational issues than standard Bayesian decision theory.

optimal solution that cannot be computed rapidly enough to apply to a real situation (c.f., Gigerenzer et al., 1999).

There has been a vigorous debate over how to interpret the probabilities that appear in a decision model (e.g., Howson and Urbach, 1993; Fine, 1973). The dominant view in artificial intelligence is subjectivism, in which probability is viewed as a measure of the degree of belief of a rational agent about uncertain hypotheses. Subjectivists assign probabilities to any hypotheses about which they are uncertain. To a subjectivist, reasonable individuals may assign different probabilities to the same outcome. The only requirements are that beliefs must conform to the mathematical constraints of the probability calculus and may not contradict evidence known to the agent. Although the subjectivist view is gaining favor in statistics, until recently the frequentist view has dominated. To a frequentist, probabilities are limiting frequencies for long sequences of outcomes generated by intrinsically stochastic systems. Unlike subjectivists, frequentists regard it as illegitimate to assign probabilities to individual events or to assign non-trivial probabilities to hypotheses with definite but unknown truth-values. Probabilities apply only to chance set-ups that produce repeatable sequences of random events. Frequentists view probability as an objective property of a chance set-up. If two individuals assign different probability distributions to a chance set-up, at least one of them must be wrong. Another interpretation of probability is the propensity view, in which probabilities refer to tendencies for events to occur under specified conditions. Unlike subjectivist probabilities, propensities are considered to be objective properties of the process that generates the outcome. Unlike freqentist probability, propensities can apply to individual events. David Lewis (1980) argued that the propensity and subjectivist views of probability can be reconciled, and that subjective probabilities should agree with objective propensities when the propensities are known.

Although subjectivists are sometimes criticized for lack of objectivity, the ability to represent and reason with subjective information is a necessary aspect of intelligent behavior. Furthermore, some of the most successful applications of probability theory (including classical statistical mechanics) are to problems involving incomplete knowledge about a system whose underlying dynamic is assumed to be deterministic. Many varieties of objectivist philosophy regard zero and one as the only allowable assignments to statements of determinate but unknown truth-value (such as the current state of a system evolving deterministically from a definite initial state). Subjectivists can perform plausible reasoning about such systems, a capability that is often needed in practical applications. Another strength of the subjectivist approach is its ability to handle small data sets and large numbers of parameters, a situation that occurs frequently in the types of problems encountered in artificial intelligence and machine learning. The theory of precise measurement (DeGroot, 1970; von Winterfeldt and Edwards, 1986) identifies conditions under which subjectivist agents beginning with different prior probabilities will converge to nearly identical posterior probabilities. The conditions under which these results hold are also characteristic of situations that might be governed by propensity and for which

limiting frequencies might exist. Thus, the frequentist, propensity and subjectivist views can be reconciled on many kinds of problems, but subjectivists extend the domain of applicability of probability beyond what frequentists or propensity advocates consider legitimate. Many applications of probability in machine learning, data mining and artificial intelligence fall outside the zone of applicability of frequency or propensity theory, and a subjectivist interpretation is required.

Subjectivist Bayesian decision theory demands conformance to rationality principles that are empirically questionable and may be computationally unachievable. Recently developed game-theoretic interpretations (Dawid and Vovk, 1999; Nau and McCardle, 1991; Shafer and Vovk, 2001) regard probability as arising out of the behavior of interacting agents, not necessarily Bayesian, who receive rewards for correctly forecasting events. In a game-theoretic setup, agents participate in an economic system in which they can announce forecasts, make bets, and/or buy and sell contingent options whose values depend on the outcomes of uncertain events. If the market is sufficiently liquid and the rules of interaction permit opportunities for arbitrage[2] to be exploited, then consistent probability forecasts can be expected to emerge from the prices at which contingent options are traded. De Finetti (1974–75) showed that any agent who violates the axioms of decision theory would agree to a sequence of transactions resulting in a sure loss. Agents violating the rationality axioms thus present arbitrage opportunities that can be exploited by other agents, and if not corrected, will lead to bankruptcy. Agents remaining in the market are driven by market pressure to behave as approximate utility maximizers. Prices for contingent options in such a market can be viewed as market consensus probabilities for the contingencies on which the options depend. There is evidence that markets for contingent options can provide more accurate probability estimates than standard methods of eliciting probabilities from experts (Berg et al., 2001). Unlike standard axiomatic decision theory, market-based evolutionary theories do not impose rationality axioms as constraints. Rather, selective pressure for rationality is only one of the "forces" operating on agents.[3] For this reason, its advocates argue that game-theoretic probability is a more satisfying foundation for probabilistic knowledge representation than its competitors.

There are numerous other philosophical positions on the types of problem to which probability may legitimately be applied, and on the meaning of the numerical

---

[2] Arbitrage means executing a sequence of trades leading to a riskless profit. Efficient markets evolve prices that eliminate opportunities for arbitrage.

[3] Fienberg and deGroot (1982) showed that proper scoring rules (rules that reward correct probability assessments) can be decomposed into components measuring *coherence* (conformance to the laws of probability), *calibration* (fit to empirical frequencies), and *refinement* (the ability to make fine distinctions). While it is true that an incoherent agent can always improve its score by finding and eliminating inconsistencies, it may be the case that the potential for improvement by improving calibration or refinement is far greater. When resource costs are taken into account and the goal is global task performance rather than accurate probability forecasts, it may be optimal to sacrifice strict coherence for better performance. See also the discussions of rationality in Russell and Norvig (2002).

probabilities thus employed (e.g., Williamson (2004) gives a list of interpretations that partially overlaps the interpretations discussed here). In particular, there is considerable debate over whether objective probabilities exist as an aspect of the physical world, and if so, to what kinds of phenomena they can be assigned. Many subjectivists insist that there is no such thing as an objective probability. Nevertheless, a theorem due to de Finetti provides a bridge for communication across the subjectivist/objectivist divide. De Finetti showed that if a subjectivist believes a series of events is exchangeable, then her belief assignments will be identical to those of an objectivist Bayesian who believes the events are independent and identically distributed trials of an objectively stochastic system, and who assigns a given prior distribution to the unknown probability. That is, the subjectivist and objectivist agree on a mathematical model for the empirical phenomenon, but differ in the meaning they attach to the mathematical constructs. Furthermore, there is no conceivable empirical evidence that would discriminate between the subjectivist and objectivist metaphysical viewpoints. Inaccessibility to empirical test does not, of course, imply that there is no truth of the matter, but it does mean that there is no scientific basis for ascertaining the truth of the matter if it exists.

2.2. GRAPHICAL PROBABILITY AND DECISION MODELS

Explicitly representing and reasoning with all possible exceptions and contingencies results in a combinatorial explosion of possibilities. Graphical probability models have become popular because they can tractably represent and perform inference on reasonably faithful models of the uncertainties involved in realistically complex tasks. In a graphical probability model, a directed or undirected graph is used to represent qualitative information about probabilistic dependencies. Nodes in the graph represent *random variables*, or sets of mutually exclusive and collectively exhaustive hypotheses. Edges in the graph represent direct dependencies of a random variable on the value of its neighboring random variables. Quantitative information about the strength of dependency is represented by local probability distributions associated with the nodes in the graph. Whereas the resources required to store and/or compute with a general probability distribution are exponential in the number of random variables, knowledge representation in a graphical probability model with a bounded number of neighbors per node scales linearly in the number of random variables. When the graph is singly connected (i.e., there is only one path between any two random variables), inference also scales linearly with the number of random variables. Although there are special cases in which a singly connected graph is adequate, realistic tasks often require more complex connectivity. Exact inference algorithms have been developed for multiply connected graphs. Although their worst-case complexity is exponential, there are interesting classes of problems for which exact methods are tractable. In the general case, approximate inference is required. A number of general-purpose methods have been developed for approximating Bayesian inference in multiply connected dependency graphs.

We illustrate graphical models using a case study based on the popular Paramount series *Star Trek*. The setting for our case study is the Starship *Enterprise* in the 24th Century (see Figure 1). Our task is to detect Romulan starships (considered hostile by the United Federation of planets) and assess the level of danger they pose to our own starship, the *Enterprise*. Figure 2 shows a directed graphical model, or Bayesian network, for a simplified version of this task. Starship detection is performed by the *Enterprise's* suite of sensors, which can correctly detect and discriminate starships with an accuracy of 95%. However, Romulan starships may be in "cloak mode," which would make them invisible to the *Enterprise's* sensors.
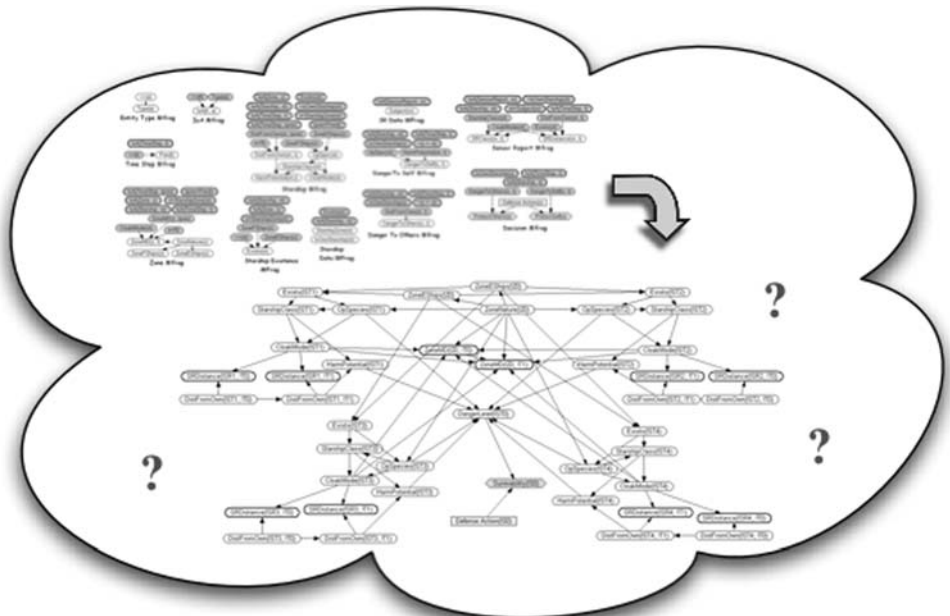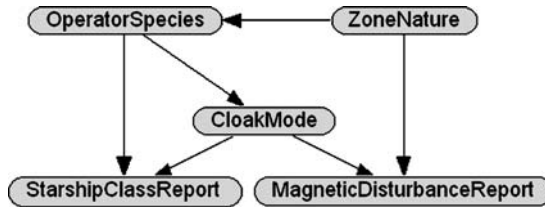

Figure 1. A 24th century decision support system.

Figure 2. Bayesian network.

Even for the most advanced sensor technology, the only hint of a nearby starship in cloak mode is a slight magnetic disturbance caused by the enormous amount of energy required for cloaking. The *Enterprise* has a magnetic disturbance sensor, but it is difficult to distinguish background magnetic disturbance from that generated by a nearby starship in cloak mode. This Bayesian network can be used to compute updated probabilities for some random variables given information on other random variables. For example, if starship class and magnetic disturbance reports were received, Bayes rule could be used to revise the probabilities for the unobserved nodes to reflect the information contained in the reports.

The reports from the friend/foe and magnetic disturbance sensors are obtained by processing the raw sensor data. Although not pictured in the figure, the relationship between the evidence random variables and the raw sensor data is also probabilistic, and can itself be represented in the language of graphical models (e.g., Binford and Levitt, 2003; Grenander, 1996). Thus, graphical models provide a consistent, theoretically justified theory of knowledge representation and evidential reasoning that spans the subsymbolic through the cognitive levels.

The example of Figure 2 can be extended to a decision graph, or influence diagram, as shown in Figure 3. Two new types of node have been introduced: *decision* nodes, represented as rectangles, and *utility* nodes, represented as hexagons. Decision nodes represent choices available to the reasoning agent. Utility nodes measure how well the agent's objectives are satisfied. The model of Figure 3 represents a
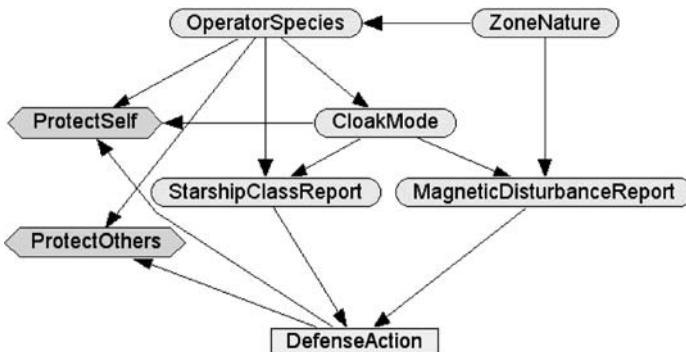

Figure 3. Decision graph.

decision of which defensive action to take (none; fire weapon; retreat). The components of utility are the danger to one's own ship and the danger to nearby friendly ships. The arcs entering a decision node represent information available to the reasoning agent at the time the decision is made. In this example, the reasoning agent knows the results of the sensor reports. The reasoning agent can choose as its policy any function of the information available at decision time. The optimal policy maximizes the sum of the mathematical expectations of the utility nodes given the available information. In this example, harm to own or friendly ships is modeled by negative utility, or loss. Thus, the model of Figure 3 would recommend a function of the sensor reports that minimizes the total expected loss to own and friendly ships due to enemy fire plus the cost of any harm caused by mistakenly firing on friendly ships.

A natural question to ask about Figure 3 is why, if a decision graph represents a mathematical model for value-driven decision making, there is no arc from the utility nodes into the decision node. A naïve reading of the graph might give the impression that the agent's choice is not affected by the potential loss from enemy fire and fratricide. To understand why this is not the case, it is necessary to examine the semantics of the diagram more closely. The diagram and corresponding numerical information specify a mathematical model of task-relevant aspects of the world viewed from the reasoning agent's perspective. The arcs represent three kinds of influences. Arcs into world state nodes represent deterministic or stochastic relationships. For cause and effect relationships, the convention is to draw the arc from cause to effect; for correlations, the arc can go in either direction. Arcs from world state nodes into decision nodes reflect information available to the agent at the time of choice. Arcs into utility nodes represent mappings from situations in the world to degrees of satisfaction the agent experiences on the corresponding dimensions of value. Arcs may not exit utility nodes. The agent's experience of satisfaction occurs *as a result* of the choice, and *at a later time* than the agent's action. Thus, the actual experienced satisfaction can neither cause the agent's action, nor be available as information on which the agent can base the decision.

The effect of values on decisions is brought about when the agent solves the decision graph for the optimal policy, and then uses the solution to decide which action to take. Conceptually, solving the graph proceeds in two steps. The first step is to compute the mathematical expectation of the total utility for each of the choices given each of the possible information states of the reasoning agent at the moment of choice. This expectation is a function mapping values of *TypeReport*, *MDReport*, and *DefenseAction* to real numbers representing the total expected loss if the reasoning agent observes the given location and configuration and takes the indicated action. The second step in solving the graph is to select an optimal decision policy that maps information states to actions. The optimal policy maps an information state (value of *TypeReport* and *MDReport*) to the action for which the expected loss is lowest. After the diagram has been solved, the optimal policy is stored with the *DefenseAction* node. The agent then receives observations on

location and configuration, looks up the optimal action corresponding to these observations, and executes that action.

The graph of Figure 3 is simple and easily solved. Decision graphs for realistically complex problems can easily become intractable. There is an extensive and rapidly growing literature on exploiting independence relationships among random variables and decomposability of value functions to perform efficient computation of optimal or approximately optimal decision policies (e.g, Jensen, 2001; Neapolitan, 2003; Boutilier et al., 1999).

The model of Figure 3 was constructed from the point of view of the reasoning agent. Although this model does not explicitly represent the decision problem of other agents, it is clear that the commanders of other starships face similar decision problems. To the reasoning agent, the utility and decision nodes of the other starships would be represented as world state nodes. From the point of view of one of the other starships, the utility and decision nodes of the *Enterprise* would be represented as world state nodes. From the point of view of an external observer, all nodes would be uncertain world state nodes. Formulating and solving a model of another agent's decision problem can be a useful aid to predicting the other agent's behavior, because the agent can be expected to select actions that serve his or her objectives given the information he or she has available, taking cognitive limitations into account.

2.3. MULTI-ENTITY BAYESIAN NETWORKS AND DECISION GRAPHS

Standard Bayesian networks and decision graphs are limited to problems with a fixed number of uncertain hypotheses in which all the relevant variables and relationships can be specified in advance of problem solving. This restriction is inadequate for complex real-world problems involving an unspecified number of objects of different types interacting in varied ways. There are many questions of interest to the *Enterprise* and its crew that demand greater expressive power than standard Bayesian networks can offer. As an example, we cannot know in advance how many starships the *Enterprise* is going to encounter. Even if we were to build a Bayesian network for each possible number of nearby starships, if the number of nearby starships is uncertain, we would not know which one to use. We also cannot specify in advance the relationships among nearby ships, e.g., whether they are isolated ships operating independently or are acting as a group. In short, Bayesian networks lack the expressive power to represent entity types (e.g., starships) that can be instantiated as many times as required for the situation at hand, and can be related to each other in varied ways (e.g., operate in groups).

Another well-known limitation of standard Bayesian networks is their lack of support for recursion. For example, the magnetic disturbance caused by a starship in cloak mode would show a characteristic temporal pattern. Standard Bayesian networks do not provide a natural way to represent such repeated patterns. Dynamic Bayesian networks (Murphy and Russell, 2000) and partially dynamic Bayesian

networks (e.g. Takikawa et al., 2001) extend Bayesian networks to model temporal patterns. However, there is no standard means to represent general recursive probabilistic relationships.

A rapidly growing research area is the development of extensions to the language of graphical models to enable representation of various kinds of repeated substructures. For example, if there were many starships nearby, an *OperatorSpecies* node would be needed for each of them. If they were operating independently, these nodes would be independent copies of the node appearing in Figure 2; if they were operating as a group, they would be correlated because one would expect the starships in a coordinated group to be operated by allied species. For the *Starship-ClassReport* node, many instances would be needed per starship, each representing a report received at a different time.

A number of recently developed languages extend graphical models to represent repeated sub-structures and recursive relationships. Examples include pattern theory (Grenander, 1996), hidden Markov models (Elliott et al., 1995), the plates language implemented in BUGS (Gilks et al., 1994; Buntine, 1994; Spiegelhalter et al., 1996), object-oriented Bayesian networks (Koller and Pfeffer, 1997; Bangsøand Wuillemin, 2000; Langseth and Nielsen, 2003), probabilistic relational models (Getoor et al., 2001; Pfeffer, 2000), and multi-entity Bayesian networks and decision graphs (Laskey, 2005; Laskey and Costa, 2005). Decision graphs can also be extended to multi-agent problems, in which each agent has its own utility and decision nodes, and each agent's optimal policy is to maximize the expectation of its utility nodes conditional on its available information (e.g., Kearns and Mansour, 2002). Attractive features of graphical models as a language for representing knowledge are their principled treatment of uncertainty, their provision for specifying knowledge as modular components with well-defined interfaces, and the existence of general-purpose exact and approximate inference and learning algorithms.

Figure 4 shows a more complex version the *Star Trek* model expressed in the language of multi-entity decision graphs. Multi-entity Bayesian network (MEBN, pronounced "mee-ben") logic is a formal system that combines the expressive power of first-order logic with a sound and logically consistent treatment of uncertainty (Laskey, 2005; Laskey and Costa, 2005). MEBN provides syntax, a set of model construction and inference processes, and semantics that together provide a means of defining and reasoning with probability distributions over unbounded and possibly infinite numbers of interrelated hypotheses. MEBN can express a probability distribution over models of any consistent, finitely axiomatizable first-order theory. Multi-entity decision graphs (MEDG, pronounced "medge") extend MEBN to include decision and utility nodes.

Knowledge about attributes of entities and relationships among entities is expressed as a collection of MEBN fragments (MFrags) organized into MEBN Theories. An MFrag contains a set of nodes connected by directed arcs. The nodes represent random variables related to each other by conditional dependence relationships. Each node is labeled with an expression consisting of a random variable
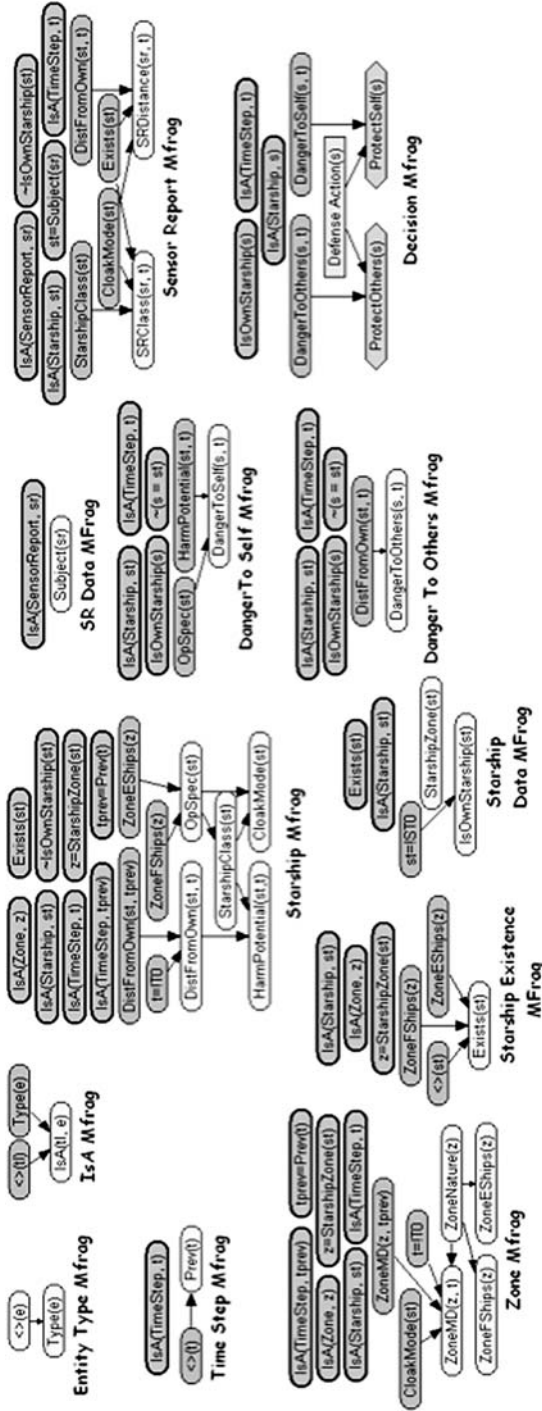
Figure 4. Multi-entity decision graph.

name and a list of zero or more arguments. This expression is a template for constructing instances of the associated random variable. Arguments beginning with lowercase letters are called ordinary variables, and serve as placeholders for actual entities. Instances of the random variable templates are constructed by replacing the ordinary variables with entity identifiers. Nodes are divided into *resident* nodes, which represent random variables whose distribution is defined in the MFrag, *input* nodes, which condition the distribution of the resident nodes, and *context* nodes, which represent Boolean conditions that must be satisfied for the dependency relations and local distributions to apply. An MFrag specifies a conditional probability distribution for instances of its resident random variables given their parents in the fragment graph and the context nodes. A MEBN Theory is a set of MFrags that collectively satisfies consistency constraints ensuring the existence of a unique joint probability distribution over instances of the random variables represented in its MFrags. A MEDG theory consists of a collection of partially specified graphical models, called MEDG fragments, that collectively specify a decision model involving a possibly uncertain and possibly unbounded number of interacting entities.

Each MFrag represents a small, separable component of knowledge that can be instantiated as many times as required in a given situation. For example, the random variable *HarmPotential*(*st*, *t*) is resident in the *Starship* MFrag, and represents the potential for harm to own starship from a starship *st* at time step *t*. To refer to the harm potential from an actual starship at an actual point in time, unique identifiers are substituted for the arguments *st* and *t*, respectively. MEBN logic contains built-in MFrags for function composition, logical connectives, and quantifiers. There is also a special *identity* random variable $\diamond(e)$, which maps its argument to the unique identifier for the entity it represents, if it represents an actual entity, or to the special value $\perp$ (meaning *absurd*) if there is no entity it represents. The special value $\perp$ allows MEBN logic to represent both finite and infinite domains (there are infinitely many unique identifiers, but in a finite domain, all but finitely many have value $\perp$). Indeed, MEBN theories can represent uncertainty over whether the domain is finite or infinite.

A domain model is constructed by augmenting the built-in logical MFrags with a set of modeler-defined domain-specific MFrags. Figure 4 contains twelve domain-specific MFrags for the Starship MEBN theory.[4] A domain-specific MFrag specifies modeler-defined structure and local distributions for a set of random variables used to model a domain. Local distributions are defined using a local expression language with first-order expressive power, and map configurations of the parents of a random variable to probability distributions for the possible values of the random variable.

---

[4] A full specification of the Starship MEBN theory can be downloaded from http://www.pr-owl.org. The files include executable code for performing inference using the Quiddity*Suite relational probabilistic modeling toolkit. The site also contains a version written in PR-OWL, a probabilistic extension to the OWL ontology language (Costa, 2005).

The *Entity Type* and *IsA* MFrags provide the logical machinery necessary for reasoning about different types of entities. The *Entity Type* MFrag defines the types of entity that exist in the Starship domain. In a typed MEBN theory, unique identifiers are partitioned into type-specific subsets. *Type(e)* maps its argument to a type label for the type of entity it represents. The *IsA* MFrag defines the Boolean random variable *IsA(tl, e)*, which has value T (true) when its second argument is an instance of the type represented by its first argument. In Section 4 below, we will consider a more complex type system with subtyping, inheritance, and polymorphism.

The *TimeStep* MFrag represents temporal recursion. Recursive relationships can be represented in MEBN logic, provided that the context constraints on the MFrags do not allow a random variable instance to be an ancestor of itself. Given a total ordering of the time step unique identifiers with unique fixed point !*T0*, the local distribution for *Prev(t)* maps !*T0* to itself, and each time step other than !*T0* to its predecessor in the total ordering.

The *Starship*, *Starship Existence* and *Starship Data* MFrags represent knowledge about the properties of starships and their interactions with other entities. The *Starship Data* MFrag defines the unique identifier used to refer to our own starship and defines a prior distribution for the zone where the action takes place. The *Starship* MFrag defines distributions for the species operating a starship, its likely distance from our own starship, its starship class, whether it is in cloak mode, and its potential to harm our own starship. The *Starship Existence* MFrag defines a distribution for the random variable *Exists(st)*. This Boolean random variable is used to reason about whether or not its argument refers to an actual starship. This random variable allows us to hypothesize starships to explain possibly spurious sensor reports, or to express uncertainty about the number of starships in the zone.

The *Zone* MFrag represents knowledge about the zone in space. *ZoneNature* represents whether the action takes place in deep space, near a planetary system, or near the boundary of a black hole. *ZoneEShips* and *ZoneFShips* represent the number of friendly and enemy ships in the zone. *ZoneMD* represents the magnetic disturbance in the zone at a given time. Instances will be independent random variables if no active cloaking is activated and will exhibit a characteristic temporal fluctuation in the presence of one or more cloaked starships.

The *Sensor Report* and *SR Data* MFrags represent the connection between reports and the starships that generated them. *Subject(sr)* represents the subject of a report; *SRClass(sr, t)* and *SRDistance(sr, t)* represent the reported class of the subject and its reported distance from own starship, respectively. Note that in this simple model, we are assuming that each report represents a time series of observations on a given starship. In a more complex model with uncertainty about the association of reports to objects, the subject of a report would also have to be indexed by time.

The *DangerToSelf* and *DangerToOthers* MFrags define probability distributions for the level of danger to our own and friendly starships. The *Decision* MFrag defines the utility nodes and the options for defensive action.

The MFrags of Figure 4 represent a generative theory of the Starship domain. That is, they define a joint probability distribution over situations involving different numbers of actual starships, zones, time steps and reports. To reason about actual situations, the generative theory is augmented with findings that specify the values of particular random variable instances. Next, a set of target random variables is specified for a query. Finally, a *situation-specific Bayesian network* (SSBN) is constructed to calculate the posterior distribution of the target random variables given the findings (Figure 5). In some cases, the response to the query can be computed using a finite Bayesian network. In other cases, the answer can be obtained as the limit of a sequence of situation-specific Bayesian networks of increasing size. It may be possible to calculate results for some parts of some queries without explicitly constructing all computationally relevant nodes (c.f. Poole, 2003; Mahoney and Laskey, 1998).

2.4. MEBN Semantics

In the standard semantics for first-order logic developed by Tarski (1944), a first-order theory is interpreted in a domain by assigning each constant symbol to an element of the domain, each function symbol on k arguments to a function mapping k-tuples of domain elements to domain elements, and each predicate symbol on k arguments to a subset of k-tuples of domain elements corresponding to the entities for which the predicate is true (or, equivalently, to a function mapping k-tuples of domain elements to truth-values). If the axioms are consistent, this can be done in such a way that all the axioms of the theory are true assertions about the domain, given the correspondences defined by the interpretation. Such an interpretation is called a model for the axioms.

MEBN theories define probability distributions over models of an associated first-order theory (Laskey, 2005). Entities are referred to by a countable set of *unique identifiers*. Instances of non-Boolean random variables specify random functions from the entity identifiers into the entity identifiers. Instances of Boolean random variables specify random functions from entity identifiers to truth-values. Non-Boolean and Boolean random variables correspond to functions and predicates in classical first-order logic. The MFrags specify probability distributions for the values of functions and predicates.

A MEBN theory is interpreted in a domain of application by associating each entity identifier symbol with an entity in the domain. Through this correspondence between identifiers and the entities they represent, the probability distribution on entity identifiers induces a probability distribution on attributes of and relationships among entities in the domain of application. In particular, although the generative distribution for a MEBN theory constructs interpretations in the countable domain of entity identifiers, a MEBN theory can be applied to reason about domains of any cardinality. Although at most a countable infinity of actual entities can be labeled by unique identifier symbols, the domain itself can be of any cardinality.
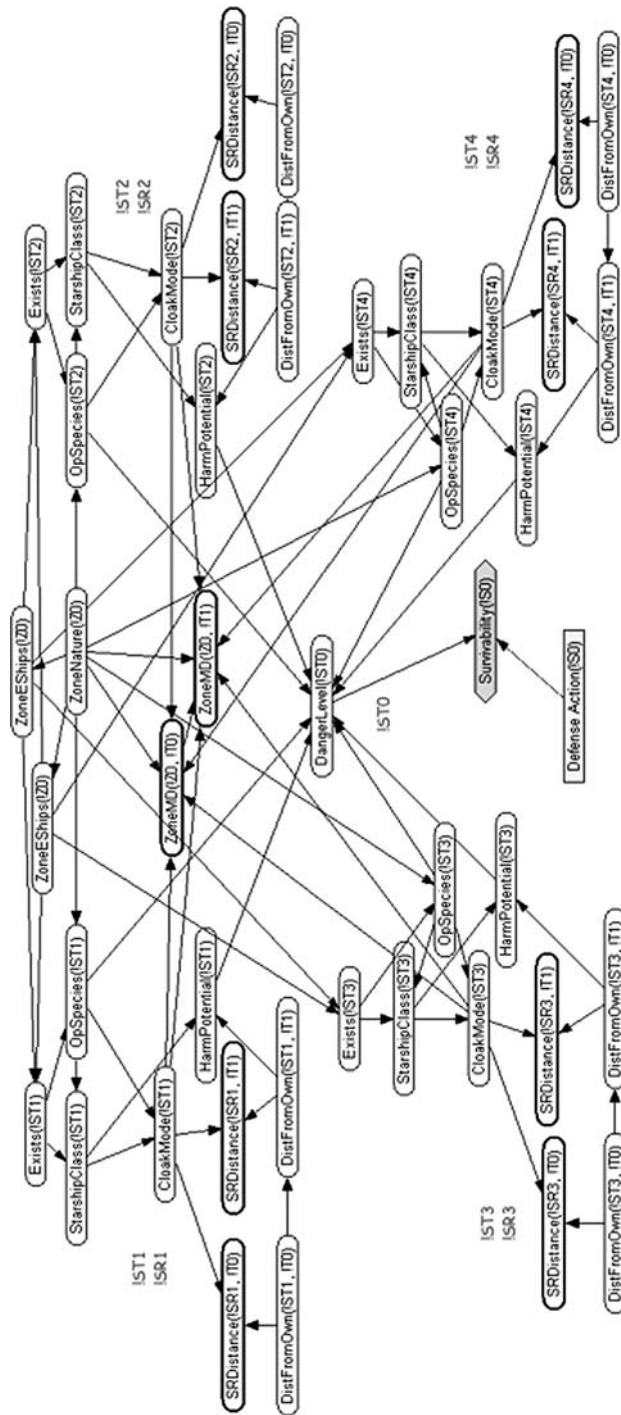
Figure 5. Situation-specific bayesian network.

Advantages of the MEBN random variable semantics are clarity and modularity. The mathematical function represented by a given random variable does not change when changes are made to unrelated parts of the representation. For example, we could add a new collection of MFrags to our theory, say for reasoning about the dietary habits of crew members, without affecting the probabilities of any assertions unrelated to the change. Furthermore, the probability distribution represented by a MEBN theory is a well-defined mathematical object independent of its correspondence with actual objects in the world, having a clearly specified semantics as a probability distribution on entity identifiers and truth-values, regardless of the meaning assigned to the entity identifiers. The adequacy of a MEBN theory for reasoning about the actual world rests in how well the relationships in the model reflect the empirical relationships among the entities to which the symbols refer in a given domain of application. Our approach thus enforces a distinction between logical and empirical aspects of a representation and provides a clearly defined interface between the two.

2.5. REASONING ABOUT CAUSE AND EFFECT

Reasoning about causality is fundamental to intelligence. In our example, the *Enterprise* risks destruction if it takes no defensive action against a hostile starship, and might cause great harm if it fires upon a friendly starship. Captain Picard must predict the likely effects of each option, and balance these risks when deciding which course of action to take. His life and the lives of his crew depend on his ability to evaluate empirical evidence to draw inferences about cause and effect.

Causal claims are stronger than claims about correlation (Pearl, 2000). A causal claim asserts not only that the values of two random variables are correlated, but also that the association is stable under interventions that do not disturb the causal connection. For example, the statement that cloaking devices cause a magnetic disturbance implies not just that we expect greater magnetic disturbance when there are cloaked starships nearby, but also that activating (or deactivating) a starship's cloaking device is likely to increase (or decrease) the magnetic disturbance in the surrounding region of space.

Although Bayesian networks with different graphical structures can represent the same joint probability distribution, when relationships are causal, it is conventional to orient the arcs in the direction of causation (cf., Druzdzel and Simon, 1993; Pearl, 2000). This convention tends to result in more parsimonious and intuitively natural models, which simplifies knowledge engineering and learning. Directed graphs are a natural way to represent *intercausal dependence*, whereby two *a priori* independent events become correlated when a common effect is observed. For example, suppose a team that was beamed down to a planet's surface is several hours late in reporting back to the *Enterprise*. Captain Picard's concern about their safety will be greatly alleviated if he learns that today is the most solemn

holiday of the planetary religion, and therefore, no off-planet communications are permitted. Evidence of the communication blackout *explains away* the missing status report. Although harm to the crew would be a probable explanation if today were not a religious holiday, after learning about the holiday, the harm hypothesis has become improbable because it is no longer needed to explain the evidence. The two hypotheses, harm to the crew and a religiously motivated communications blackout, were *a priori* independent but have become dependent due to the missing status report.

Another important advantage of using Bayesian networks to model causality is their use in predicting the effect of natural or deliberate changes in the value of a random variable. We can predict, for example, that turning on a cloaking device will prevent sensors from detecting the cloaked ship, and is likely to give rise to an increase in the magnetic disturbance. On the other hand, correlations due to intercausal dependence are evidential, a property of an agent's incomplete knowledge of a pre-existing situation. Thus, it would do no good to a landing party being held hostage to declare a religious holiday and impose a moratorium on communications. In other words, *evidence* of the religious holiday affects Captain Picard's *belief* that his landing party is safe, but intervening to *cause* a religious holiday has no effect on the landing party's *actual* safety. Causal links in a Bayesian network can be used to predict which interventions might change a given random variable and which cannot (Pearl, 2000).

Formal logical and mathematical tools for analyzing statistical associations have attained a high degree of sophistication. Until recently, however, formal tools for reasoning about causal relationships have received much less attention. By providing a formal mathematics for expressing and reasoning about cause and effect relationships, directed graphical models provide a formal basis for a science of cause and effect relationships. Pearl says:

> This mathematical language is not simply a heuristic mnemonic for displaying algebraic relationships. . . Rather, graphs provide a fundamental notational system for concepts and relationships that are not easily expressed in the standard mathematical languages of algebraic relationships and probability calculus. Moreover, graphical methods now provide a powerful symbolic machinery for deriving the consequences of causal assumptions when such assumptions are combined with statistical data.

In summary, directed graphical models provide a formal system for expressing theories of cause and effect relationships, deriving the consequences of causal theories, evaluating alternative causal theories in the light of evidence, and assessing the degree to which competing causal claims are empirically distinguishable. All branches of science can benefit from a scientifically principled methodology for evaluating causal claims and assessing their degree of empirical confirmation. Better formal tools for analyzing causality are especially relevant to quantum theory, a

fundamental scientific theory thought to encompass all natural phenomena, that has been plagued since its inception with bitter disputes over interpretation.

3. Quantum Theory and Bayesian Decision Theory

Classical mechanics is a dynamically complete theory with no role for agency, knowledge or efficacious deliberate choice. Once initial conditions are specified, a classical physical system follows a definite trajectory that, at least in principle, can be predicted with absolute precision indefinitely into the future. Of course, in practice this predictability is limited by approximation and measurement error in the specification of both the initial conditions and the parameters of the dynamical equations. Nevertheless, in principle, the evolution of a classical physical system is perfectly determined by initial conditions. Furthermore, classical mechanics has nothing at all to say about the processes by which intelligent agents make predictions about the behavior of physical systems or make and implement decisions.

Early in the 20th century it was discovered that the classical picture of a world of perfectly deterministic physical systems was incorrect. The classical picture was replaced by the explicitly probabilistic quantum theory. The degree of accord between the theoretical predictions of quantum theory and empirical measurements performed on quantum systems is striking. Nevertheless, many physicists remain uncomfortable with quantum theory. There are three major reasons for this discomfort. First, quantum theory makes only probabilistic predictions about the trajectory of a system. Many scientists are uncomfortable with a picture of Nature that has an intrinsically stochastic component. Second, the theory is non-local. That is, there are correlations between spacelike separated events that cannot be explained by a hidden variable theory with strictly local influences. Third, the theory contains a major explanatory gap known as the "measurement problem," in which deterministic evolution of the wave function is interrupted by "reduction events" for which current physics has no theory.

3.1. QUANTUM DYNAMICS AND MEASUREMENT

The state of a quantum system at any time is described by a mathematical structure called the *quantum state*. In the usual formulation, states are represented as density operators on a Hilbert space associated with the system.[5] A Hilbert space is a complex vector space equipped with an inner product. A density operator is a positive operator with trace equal to one that acts on the state space of the system. The state is *pure* if its density matrix has rank equal to one. Pure states represent

---

[5] Most introductory textbooks represent pure states as vectors in a Hilbert space and mixed states as probability weighted averages of pure states. The density matrix formulation is preferred by most working physicists because it is mathematically equivalent and provides a unified treatment of pure and mixed states.

maximal knowledge about a quantum system. If the state is not pure, it is called a *mixed* state. A mixed state can be represented as a weighted sum of pure states, where the weights are non-negative and sum to one. Mixed states represent uncertainty about the state of a system. Mixed states can also represent situations in which the state of a non-isolated system is *entangled* with the state of its environment. Entanglement is a quantum phenomenon in which the state of a system cannot be fully described except in relation to its environment. When a system is entangled with its environment, the subsystem alone is in a mixed state even if the system and environment combination is in a pure state.

The time evolution of an isolated quantum system is described by a unitary transformation. With each isolated quantum system is associated a characteristic one-parameter continuous group of unitary operators $U(t)$. If ρ_0 is the initial density matrix, then the density matrix after the system evolves for t units of time is given by $\rho_t = U(t)\rho_0 U(t)^*$, where $U(t)^*$ is the adjoint of $U(t)$.

The postulate of unitary evolution breaks down when a quantum system interacts with its environment. Open quantum systems can undergo a discontinuous change called *state reduction*, or more picturesquely, *collapse of the wave function*. The orthodox interpretation of quantum theory associates reductions with measurements performed on a system by scientists. Measurement involves an interaction of a measuring apparatus with the system, in which some feature of the microscopic quantum system is "amplified" to produce a macroscopically detectable change in the measurement device.

The state of a system after measurement depends on the measurement outcome, and is different from what it would have been if no measurement had occurred. After reduction, the pre-reduction state is projected onto one of a set of orthogonal subspaces of the system's Hilbert space. Specifically, each measurement has an associated a set of projection operators $\{P_i\}$ satisfying $P_i^2 = P_i$, $P_i P_j = 0$, and $\sum_i P_i = I$, where I is the identity operator[6] If the state prior to reduction is ρ, then the state after reduction is $P_i \rho P_i / tr(P_i \rho P_i)$ with probability $tr(P_i \rho P_i)$, where $tr(\cdot)$ is the trace operator. To gain some intuition for how reduction works, we can represent the density operator ρ and the projection operators $\{P_i\}$ as matrices (possibly of infinite dimension), where the entries of the projection matrices are all zeros except for the diagonals, which contain some 1's and some zeros. The projection operator P_i maps all entries of ρ to zero unless they are in a row and a column for which the diagonal element of P_i is 1, in which case, P_i leaves the entry unchanged. That is, P_i acts as the identity on a square sub-matrix of ρ and transforms the remaining entries to zero. The probability $tr(P_i \rho P_i)$ is the sum of the diagonal elements of ρ corresponding to entries of 1 in P_i. Dividing by $tr(P_i \rho P_i)$ normalizes the state so the sum of the probabilities for subsequent

---

[6] We consider only projective measurements, because more general measurements can be obtained by composing unitary transformations with projective measurements. We also limit consideration to measurement operators with a discrete spectrum.

measurements is unity. If ρ is in the subspace onto which P_i projects, then P_i acts as the identity, and the state after reduction is ρ with probability 1. We say the sets $\{P_i\}$ and $\{Q_j\}$ are *simultaneously measurable* if $P_iQ_j = Q_jP_i$ for all i, j. Applying simultaneously measurable projection sets in sequence projects the state onto the intersection of the subspaces corresponding to the two measurement results. If projection sets are not simultaneously measurable, then applying the first followed by the second need not leave the system in the subspace corresponding to the outcome of the first measurement. That is, the second measurement may change the system to a state other than the state corresponding to the result of the first measurement.

Bohm (1951) states that although the quantum state has been called a "wave of probability," it is more accurately described as a "wave from which many related probabilities can be calculated." That is, quantum theory does not specify "the" probability of an event. It specifies a conditional probability for the outcome given that a specific reduction operator is applied at a specific time. Different reduction operators applied at different times give rise to different probability distributions for outcomes, but they all can be calculated from the evolving quantum state. Normalizing the state after reduction is mathematically equivalent to applying Bayes rule to Bohm's "wave from which many related probabilities can be calculated." That is, after a measurement has occurred, the "wave of related probabilities" is conditioned on the measurement operator that was applied, the time of application, and the actual outcome that occurred. Predictions about the future behavior of the state are revised accordingly. In experimental tests, the probability forecasts produced by this recipe have proven to be stunningly accurate.

Although there exist experimental procedures, described in classical language, for effecting measurements on various types of physical systems, there are no fundamental physical laws governing how the scientist, considered as a physical system, makes the choice of which measurement, if any, to perform, and the time at which the measurement occurs. There have been many attempts to formulate quantum theory as unitary evolution without reduction. The basic approach is to embed the observed system in a larger system that contains both the observing agent and the observed system. None of these attempts has been fully successful, and physicists disagree about whether such a reformulation is possible.

Albert (1992; p. 79) states the measurement problem as follows:

> The dynamics and the postulate of collapse are flatly in contradiction with one another. . . the postulate of collapse seems to be right about what happens when we make measurements, and the dynamics seems to be bizarrely *wrong* about what happens when we make measurements, and yet the dynamics seems to be *right* about what happens whenever we *aren't* making measurements.

In other words, to reproduce the quantitative predictions of quantum theory, it is necessary to augment unitary equation with additional rules. In the standard

interpretation, these rules correspond to initiation of reduction events by agents and selection of outcomes by Nature. Our capacity to perform measurements is taken as a given empirical and phenomenological fact. However, there is no fundamental physical theory governing the timing and possible outcomes of state reduction. Statements about which measurement is applied at what time are formulated in the everyday language of classical physics, not in the language of quantum theory used to describe the system being observed. Thus, the very means by which we are able to learn about the behavior of quantum systems, and to construct theories to predict the observable consequences of their behavior, is itself unexplained by quantum theory.

To summarize, the dynamic behavior of a quantum system depends in macroscopically observable ways on a process for which physics has no theory. Although quantum theory includes statistical laws governing Nature's choice of outcome when a measurement operator is applied, the known physical laws do not fix, *even statistically*, which measurement operators are applied under what conditions. Thus, the theory contains a contingent element. It specifies behavior of the system *given* the actions an agent external to the theory takes to observe the system. But according to the orthodox interpretation, this external agent can choose which of several distinct macroscopic effects to actualize by choosing which aspects of the system to observe. This dependence of the predictions of the theory on an aspect of reality for which there is no theory worries many physicists. However, agents with brains and bodies built out of the elements studied by atomic physics have the demonstrated capacity to perform plausible reasoning and make efficacious choices. Moreover, there are aspects of the physical architecture of the brain that make it likely that quantum mechanical effects are important in its dynamical behavior (Schwartz et al., 2005). Some scientists have hypothesized that cognitive agents act on the world by controlling the timing and selection of reduction operators in order to bring about survival and life enhancing outcomes.

3.2. A MEBN REPRESENTATION OF QUANTUM EVOLUTION

In this section, we present a set of MFrags to represent the evolution of quantum systems. The MFrags for our MTheory naturally fall into four different categories. First, there is a set of *logical MFrags* that represent the logical and mathematical machinery needed to express the domain-specific MFrags. The second group of MFrags represents the state of a quantum system. The third set of MFrags represents the action of unitary transformations. The final set of MFrags represents the action of reduction operators and the observable outcomes of measurements. Together, these MFrags represent a complete MEBN theory for reasoning about the evolution of quantum systems. Thus, computational Bayesian logic can represent and reason about the behavior of quantum systems.

3.2.1. *The Logical MFrags*

Before moving into our representation of the physical and empirical aspects of quantum theory, we begin by explicitly and formally representing the logical machinery needed to express the structure of our quantum theory MFrags. The necessary logical machinery can be defined using eight MFrags (see Figure 6). We have already encountered *Type* and *IsA* MFrags in the Starship model. For our representation of quantum theory, it is useful to define a more powerful type system with subtypes and polymorphism. This requires relaxing the requirement that each random variable have a unique home MFrag, to allow different distributions to be defined in multiple home MFrags (Costa, 2005). The *SubType* MFrag and the *ParentType* random variable in the *Type* MFrag define a type hierarchy. For the present purpose a tree-structured hierarchy is sufficient. In typed MEBN, a distribution can be defined for a type and inherited by subtypes unless overridden by a new definition for a subtype. The *VCount Initialization* MFrag specifies a random variable *VCount(tl)* for each entity *tl* of type *TypeLabel*. These "virtual counts" are used to define relative
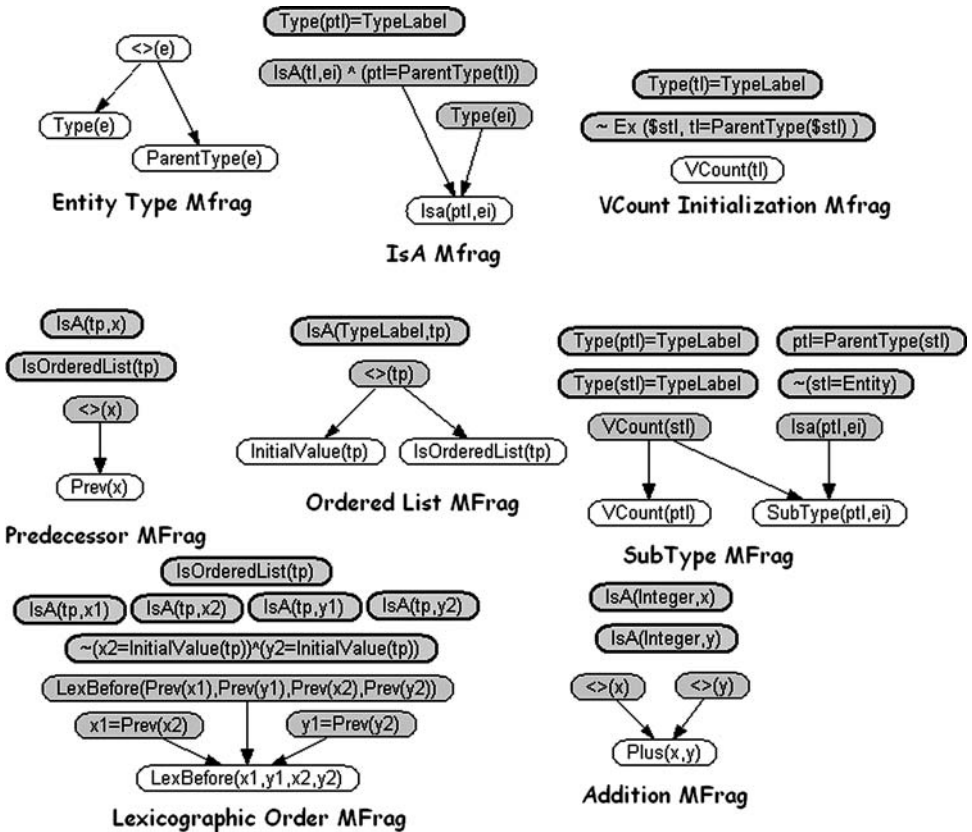

Figure 6. Logical MFrags for quantum MEBN theory.

frequencies of types. Virtual counts are needed when there is uncertainty about the types of entities. The *Ordered List* and *Predecessor* MFrags are used to reason about types that will be used as recursion indices. These include the *TimeStep* type (time steps also appear in our Starship theory), the *Integer* type, and the *Fiducial-BasisState* type, to be discussed in more depth later. To define an ordered list type, we assign value T (meaning *True*) to the *IsOrderedType* random variable, assign a unique identifier as the value of the *InitialValue* random variable, and define a function *Prev* that maps the initial value to itself and the other type-specific unique identifiers to their predecessors in the ordering. Finally, the *Addition* MFrag defines the random variable *Plus* to operate in the usual way on integers, and the *Lexico-graphic Order* MFrag defines the predicate *LexBefore*($x1$, $y1$, $x2$, $y2$) to have value T when its arguments are instances of ordered list types, and the ordered pair ($x1$, $y1$) lexicographically precedes the ordered pair ($x2$, $y2$). Obviously, other logical and mathematical functions could be defined as needed, but the ones given here are sufficient to define the remaining MFrags of our representation for quantum systems.

3.2.2. *Representing Different Types of Quantum System*

The MFrags of Figure 7 specify the types of quantum systems represented by the MEBN theory. In the *Entity Type* MFrag, each type of quantum system was assigned a unique type label, e.g., *SystemOfQubits* for a quantum computing system. The random variable *IsA*(*QuantumSystem*, *tl*) was set to have value T when the ordinary variable *tl* is replaced by the type label for a quantum system. The *IsA* MFrag also defines type labels for quantum states and quantum transformations. The three *QuantumSystemType* MFrags relate states and transformations to their respective quantum system types. The first assigns to each type of system its characteristic state and transformation types. For computational quantum systems, these
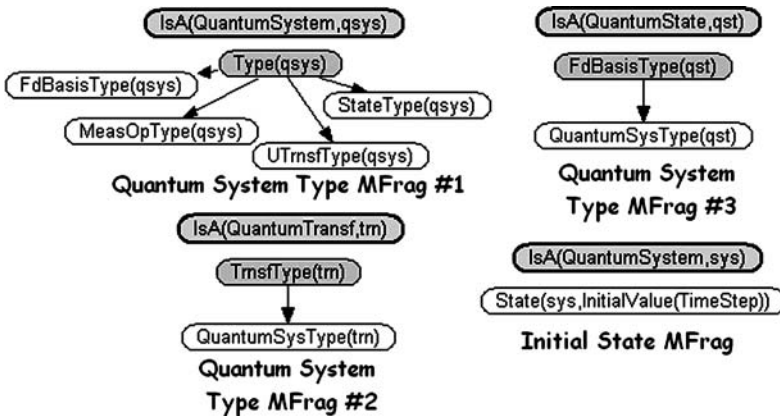

Figure 7. Quantum system type MFrags.

are *StateType(SystemOfQubits)* = *QubitVector*; *UTrnsfType(SystemOfQubits)* = *QuantumGate*; *MeasOpType(SystemOfQubits)* = *QubitVectorReduction*; and *FdBasisType(SystemOfQubits)* = *QubitConfig*. The second and third *QuantumSystemType* MFrags define the random variable *QuantumSysType*, which is a reciprocal function that maps an instance of a quantum state or quantum transformation to its type of quantum system. The fourth MFrag in this cluster specifies an initial probability distribution for a quantum system of a given type. This distribution can, of course, be modified using findings when there is situation-specific knowledge about the initial state of a system, but this MFrag provides a well-defined initial distribution when nothing is known about the system except its type.

3.2.3. *Representing the State of a Quantum System*
Our representation for states makes use of Hardy's (2001) proof that to obtain the predictions of quantum theory, it suffices to specify the action of unitary transformation and reduction operators on a discrete set of *fiducial states*. For finite dimensional systems, the required number of fiducial states is equal to the square of the dimension of the system. For systems of countably infinite dimension, a countably infinite set of fiducial states is needed. This is a useful result, because it allows us to characterize completely the behavior of a quantum system using a finite or countably infinite set of random variables.

There is more than one way to specify a set of fiducial states for a given quantum system. A natural way to choose the fiducial states is first to choose a set of mutually orthogonal projection operators that spans the state space. The elements of this set are called *fiducial basis states*. Next, for each pair of fiducial basis states, we choose two linearly independent *fiducial non-basis states* lying within the subspace spanned by the pair of fiducial basis states. For the reader familiar with Dirac bra-ket notation, the fiducial basis states can be written as $|n><n|$, where the $|n>$ are mutually orthogonal unit vectors in the system's Hilbert space and $<n|$ is the adjoint of $<n|$. For the fiduical non-basis states corresponding to $|n><n|$ and $|m><m|$, we choose the linearly independent pair of states $\frac{1}{2}(|m>+|n>)(<m|+<n|)$ and $\frac{1}{2}(|m>+i|n>)(<m|-i<n|)$. With each fiducial state ϕ we associate a *fiducial measurement* consisting of the projection operator set $M_\phi = \{P_\phi, I-P_\phi\}$. If the fiducial measurement M_ϕ is applied when the system is in state ϕ, the outcome is ϕ with probability 1. That is, fiducial measurements leave their corresponding fiducial state unchanged. The fiducial basis measurements $M_{|n><n|}$ are simultaneously measurable. If we apply the $M_{|n><n|}$ in any order to the state ρ, the state after measurement will be one of the fiducial basis states, with probability $tr(|n><n|\rho|n><n|)$ for state $|n><n|$.

To gain some intuition for the meaning of the fiducial states, consider the quantum mechanical generalization of the bit, a two-dimensional quantum system called a *qubit*. In a classical computer, a bit can be in one of two states, usually denoted 0 and 1. The corresponding qubit states are the fiducial basis states

$|0><0|$ and $|1><1|$. But unlike a classical bit, a qubit is not limited to two discrete states. A qubit can also be in a *superposition*, which can be thought of as a kind of suspension between states. A single-qubit superposition is written as $(\alpha\,|\,0> + \beta\,|\,1>)(\alpha^*<0| + \beta^*<1|)$, where α and β are complex numbers such that $\alpha\alpha^* + \beta\beta^* = 1$, and α^* and β^* represent the complex conjugates of α and β, respectively. If we apply the two fiducial measurements $P_{|0><0|}$ and $P_{|1><1|}$ in sequence, the state after measurement will be $|0><0|$ with probability $\alpha\alpha^*$ and $|1><1|$ with probability $\beta\beta^*$. Although we might be tempted to think of a qubit in superposition as being in state $|0><0>|$ with probability $\alpha\alpha^*$ and $|1><1|$ with probability $\beta\beta^*$, this is not correct. Although the superposition $(\alpha\,|\,0> + \beta\,|\,1>)(\alpha^*<0| + \beta^*<1|)$ and the mixed state $(\alpha\alpha^*|0><0| + \beta\beta^*|1><1|)$ both yield $|0><0|$ with probability $\alpha\alpha^*$ and $|1><1|$ with probability $\beta\beta^*$ when a fiducial basis measurement is applied, the outcome probabilities differ for the fiducial non-basis measurements. For example, if we apply the fiducial non-basis measurement $M_{\frac{1}{2}(|0>+|1>)(<0|+<1|)}$ to a system in the superposition $(\alpha\,|\,0> + \beta\,|\,1>)(\alpha^*<0| + \beta^*<1|)$, the outcome will be $\frac{1}{2}(|0>+|1>)(<0| + <1|)$ with probability $\frac{1}{2}(\alpha + \beta)(\alpha^* + \beta^*)$ and $\frac{1}{2}(|0> - |1>)(<0| - <1|)$ with probability $\frac{1}{2}(\alpha - \beta)(\alpha^* - \beta^*)$. On the other hand, if we apply the same measurement to the mixed state $(\alpha\alpha^*|0><0| + \beta\beta^*|1><1|)$, the two outcomes will have equal probabilities.

Hardy proved that to characterize a quantum state, it is sufficient to specify the outcome probabilities given each of the fiducial measurements. This characterization is achieved by the MFrags of Figure 8. The fiducial basis states are specified in the logical MFrags shown in Figure 6, via the definition of *InitialValue*(*FiducialBasisState*) and the *Prev*(·) function acting on entities of type *FiducialBasisState*. The *Fiducial Non-Basis State* MFrag defines the fiducial non-basis states in terms of the fiducial basis states. For each pair (*fb*1, *fb*2) of distinct fiducial basis states, the MFrag defines a unique fiducial non-basis state
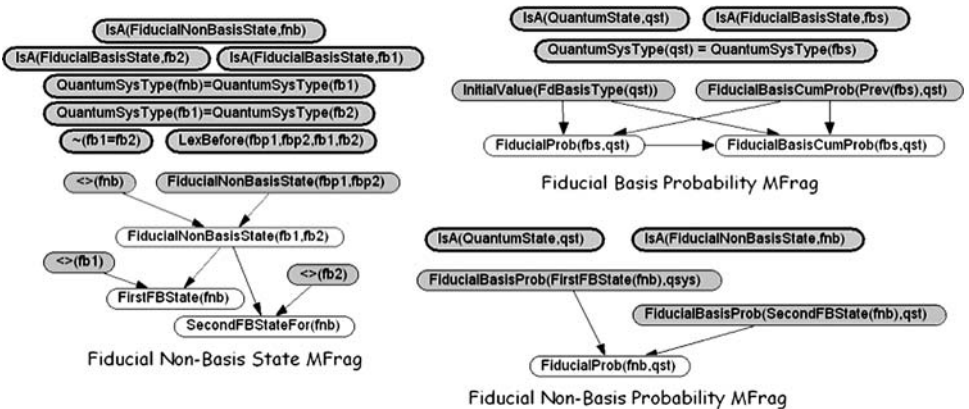

Figure 8. Fiducial state MFrags.

FiducialNonBasisState(*fb1*, *fb2*), and sets the values of *FirstFBState*(*fnb*) and *SecondFBState*(*fnb*) to *fb*1 and *fb*2, respectively. The other two MFrags of Figure 8 specify the probability distributions for outcomes when fiducial states are measured. The context constraints, parents, and local distributions ensure that the probability distributions are valid, i.e., that they respect the constraints required by the Hilbert space structure of the quantum system. Specifically, the value of *FiducialProb*(*fbs,qst*), the probability that the measurement result is *fbs* when fiducial basis state *fbs* is measured on a system in state *qst*, must be between zero and the sum of the fiducial measurement probabilities for fiducial basis states preceding *qst* in the fiducial basis state ordering. This constraint is achieved by making *FiducialBasisCumProb*(*Prev*(*fbs*), *qst*) a parent of *FiducialProb*(*fbs,qst*). Furthermore, the range of allowable probabilities for a fiducial non-basis measurement depends only on the fiducial basis measurements for the fiducial basis states of which it is a superposition. These values are therefore parents of *FiducialNonBasisProb*(*fnb, qst*).

The MFrags of Figure 8 define a generative distribution for states of each type of quantum system represented by the MEBN theory. That is, they define a joint probability distribution for situations involving different number of quantum systems of different types. The probability distributions associated with these MFrags are subjective – they represent a Bayesian agent's uncertainty about they type of system, the parameters characterizing a quantum state, etc. When there is more information about the state of an instance of a quantum system (e.g., results of measurements), that information can be applied as findings to refine the distribution for its state.

3.2.4. *Representing Unitary Transformations*

The next set of MFrags, shown in Figure 9, define unitary transformations for quantum systems. A unitary transformation transforms a system that is in state *State*(*sys*, *Prev*(*t*)) at the end of the previous time step to *PreState*(*sys*, *t*), its new state prior to any reduction operator(s) that may occur at the current time step. The *Product Factors* and *Product Transformation* MFrags are generic MFrags that define product transformations for any type of quantum system. The other two MFrags in this set define two generators for the unitary group of *SystemOfQubit* quantum systems. It is well known (c.f., Nielsen and Chuang, 2000) that any unitary transformation on a system of qubits can be obtained as the product of single-qubit transformations and CNOT gates. The MFrag for single-qubit transformations makes use of the fact that any single-qubit transformation can be expressed as global phase shift combined with a rotation (see Exercise 4.8, page 175 in Nielsen and Chuang, 2000). The parameters of this representation are represented as parents of *PreState*(*sys*, *t*) in the *Single-Qubit Gate* MFrag. Although we have explicitly shown only the unitary group generators for qubit systems, it is straightforward to extend this set of MFrags to cover any type of quantum system for which a generative distribution for the unitary group can be defined.
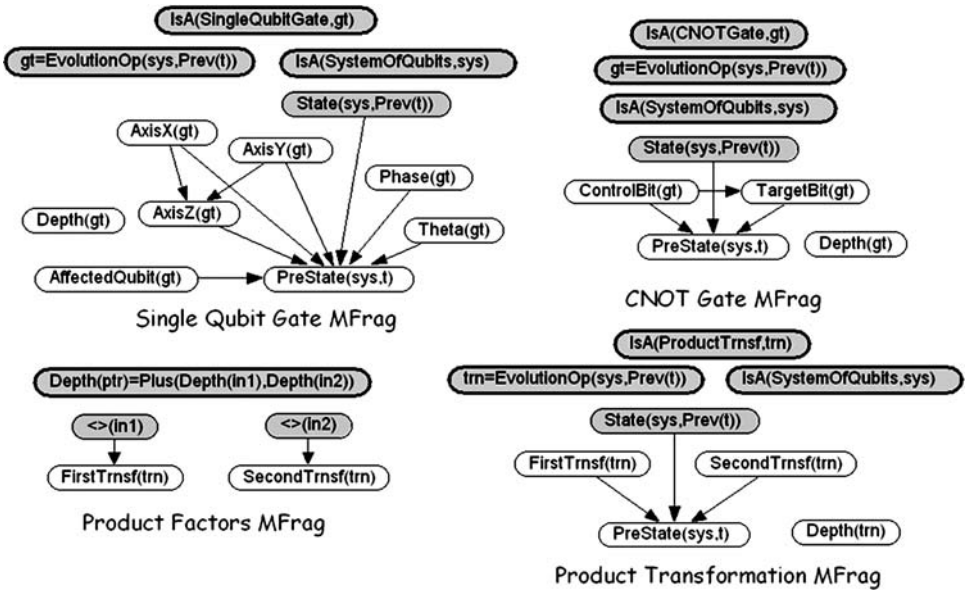
Figure 9. Unitary transformation MFrags (for Qubit Systems).

As for the preceding sets of MFrags, the probability distribution over unitary transformations defined in these MFrags represents subjective probabilities of a Bayesian agent over the unitary transformation governing the evolution of a quantum system. The distribution in these MFrags could be overridden by distributions defined for subtypes of the given type of system.

3.2.5. *Representing Reductions*

The MFrags of Figure 10 define distributions for the state after reduction events and the outcomes of measurements. Outcomes of measurements are modeled separately from post-reduction states, but are related by the restriction that the state after reduction must be in the subspace corresponding to the outcome. The *No Reduction* MFrag states simply that if no reduction occurs, then the state at the end of time step *t* is *PreState(sys, t)*, the state after applying the unitary transformation to *State(sys, Prev(t))*, and the measurement outcome is a null value indicating that no measurement was taken. The *Reduction Result* MFrag defines the distribution for the reduction result as a function of reduction operator and the fiducial probabilities. This MFrag represents only projective measurements with two possible outcomes. Because any other measurement can be constructed from two-valued projective measurements, it would be straightforward to add additional MFrags to represent other kinds of measured quantities.

These MFrags define only unitary transformations and two-valued projective measurements. The most general form of transformation for a quantum system is
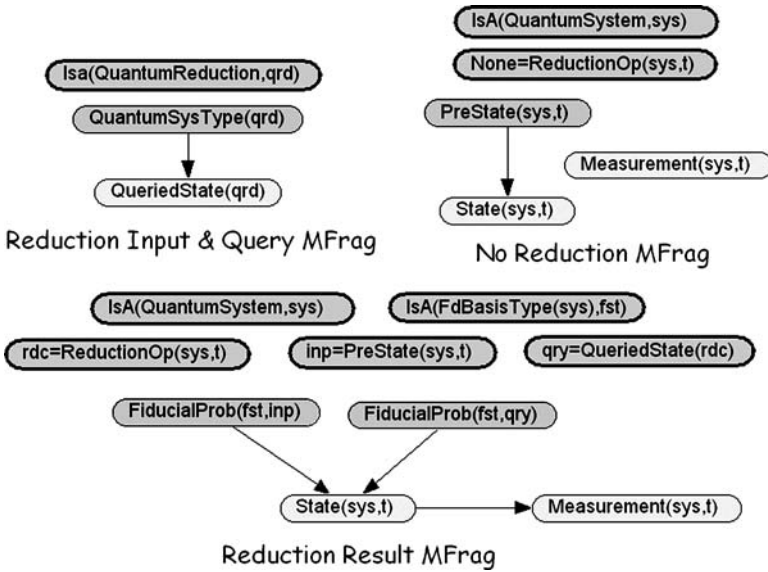
Figure 10. State reduction MFrags.

called a *quantum operation*. General quantum operations can be represented as completely positive operators on the system's Hilbert space. It can be shown that any quantum operation can be obtained as a combination of unitary transformations and reductions. Therefore, the MFrags given here are sufficient to represent the behavior of systems of qubits. Furthermore, the MEBN theory presented here can be extended to represent any quantum system for which a generative probability distribution can be defined for its unitary group.

3.2.6. *Representing the Choice of Operator*

The MFrags of Figure 6 through Figure 10 specify a MEBN theory that is complete except for the definition of two context nodes. The first of these is *EvolutionOp(sys, t)* (see Figure 9), which represents the unitary transformation that transforms the system to its next pre-reduction state. The second is *ReductionOp(sys, t)* (Figure 10), which represents the reduction operator applied at time *t*. According to standard quantum theory, a closed quantum system transforms according to a fixed, time-independent unitary transformation. Thus, for a closed quantum system there are no reductions and *EvolutionOp(sys, t)* does not depend on time. This model is depicted in Figure 11a. Open systems are subject to environmental noise. This can be represented by treating the evolution operator and reduction operators as independent realizations from the distributions represented in Figures 9 and 10, respectively, as shown in Figure 11b.

In quantum computing, the algorithm designer specifies the unitary evolution and reduction operators, and the input state is selected at run time by the user.
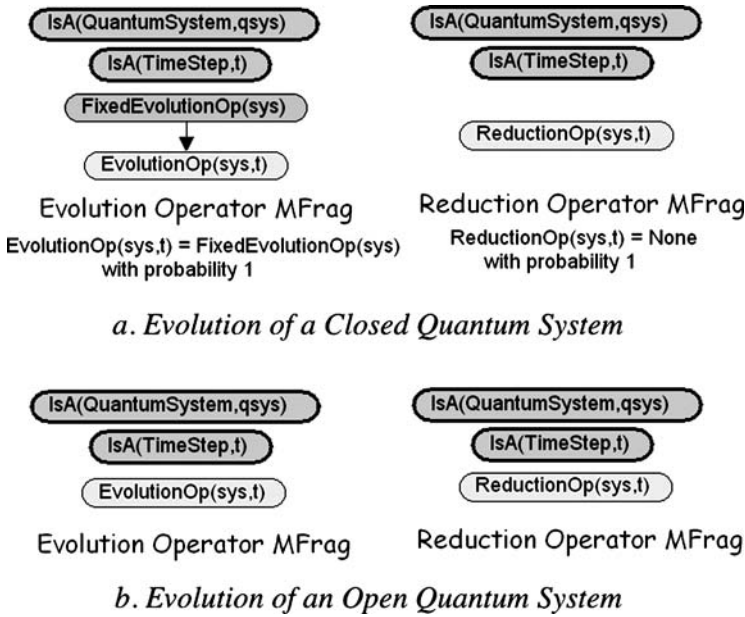
IsA(QuantumSystem,qsys)
IsA(TimeStep,t)
FixedEvolutionOp(sys)
EvolutionOp(sys,t)

Evolution Operator MFrag

EvolutionOp(sys,t) = FixedEvolutionOp(sys)
with probability 1

IsA(QuantumSystem,qsys)
IsA(TimeStep,t)

ReductionOp(sys,t)

Reduction Operator MFrag

ReductionOp(sys,t) = None
with probability 1

a. Evolution of a Closed Quantum System

IsA(QuantumSystem,qsys)
IsA(TimeStep,t)
EvolutionOp(sys,t)

Evolution Operator MFrag

IsA(QuantumSystem,qsys)
IsA(TimeStep,t)
ReductionOp(sys,t)

Reduction Operator MFrag

b. Evolution of an Open Quantum System

Figure 11. Operator generative MFrags.

The process by which these choices are made lies outside of quantum theory. These choices are treated phenomenologically by texts in quantum computing. For example:

> . . . we often speak of *applying* a unitary operator to a particular quantum system. . . Doesn't this contradict what we said earlier, about unitary operators describing the evolution of *closed* quantum systems? After all, if we are 'applying' a unitary operator, then that implies there is an external 'we' who is interaction with the quantum system, and the system is not closed. . . .for many systems like this it turns out to be possible to write down a *time-varying* [unitary evolution operator that] varies according to some parameters which are under an experimentalist's control, and which may be changed during the experiment. (Nielsen and Chuang, 2000; p. 84)

> . . . The status of [the measurement postulate] as a fundamental postulate intrigues many people. Measuring devices are quantum mechanical systems, so the quantum system being measured and the measuring device together are part of a larger, isolated, quantum mechanical system. According to [the unitary evolution postulate], the evolution of this larger isolated system can be described by a unitary evolution. Might it be possible to *derive* [the measurement postulate] as a consequence of this picture? Despite considerable investigation along these

lines, there is still disagreement between physicists about whether this is possible. We, however, are going to take the very pragmatic approach that in practice it is clear when to apply [unitary evolution] and when to apply [measurement], and not worry about deriving one postulate from the other. (Nielsen and Chuang, 2000; p. 85)

It is assumed [as one of the key elements of the quantum circuit model of computation] that any computational basis state [for an n-dimensional system of qubits] can be prepared in at most n steps. (Nielsen and Chuang, 2000; p. 202)

We argue that expressing a theory of quantum evolution in the language of Bayesian networks can help to elucidate the content of different ontological stances regarding the phenomenology, as well as to evaluate whether and to what degree they differ in empirical content. The next section presents several prominent ontologies for quantum theory and discusses them in relation to the MEBN representation of quantum evolution. With respect to our MEBN theory of quantum evolution, different ontological stances can be distinguished by differences in how the distributions for *EvolutionOp*(*sys*, *t*) and *ReductionOp*(*sys*, *t*) are defined and interpreted.

3.3. ONTOLOGIES FOR QUANTUM THEORY

The orthodox interpretation for quantum theory is associated with Bohr (1934) and is called the Copenhagen interpretation. According to the Copenhagen interpretation, quantum theory replaces a classical theory that refers to an external material universe with a new theory that refers only to the experience of observers and not to the external universe itself. Conditional on a choice of experimental set-up that defines the macroscopically detectable possibilities available to the system, quantum theory predicts the probability that each of these classically describable possibilities will occur. Proponents of the Copenhagen interpretation make no ontological commitments regarding the entities that give rise to the experienced sequence of observations. It is sometimes asserted that it is meaningless to speak of the "actual state" of a quantum system. The quantum state is asserted to be nothing but a mathematical construct for organizing the experiences of observers and enabling the computation of accurate predictions of the outcomes of experiments. According to the orthodox interpretation, quantum theory represents a set of computational rules by which scientists can make predictions about which classically describable outcomes will occur as a result of the classically describable experiments they conduct. The quantum state is a mathematical construct used to make predictions about observables, but being inaccessible to direct observation, is not to be regarded as corresponding to any definite phenomenon in Nature. The Copenhagen position is reminiscent of the strict subjectivist view that probabilities refer only to the beliefs of rational agents, and it is meaningless to speak of objective propensities. In this view, instances of the random variable template *Measurement*(*sys*, *t*) correspond to

actual experiences of an observer when a measurement is taken on a given quantum system at a given time. On the other hand, instances of *State*(*sys*, *t*) represent nothing but mathematical constructs used to compute probabilities for the different outcomes of *Measurement*(*sys*, *t*).

Although the Copenhagen interpretation is the standard view, most physicists prefer, at least informally, to operate with an ontology that connects the terms in the theory to an underlying physical reality that gives rise to the experience of observers. One important characteristic that distinguishes ontologies is whether reductions are treated as real phenomena in Nature or as artifacts of incomplete knowledge about open quantum systems. As noted above, if a system is entangled with its environment, the system considered alone will in general be in a mixed state even if the combination of system plus environment is in a pure state. Furthermore, the process of *decoherence*, in which an open system loses its quantum properties due to environmental influences, operates extremely rapidly. Some physicists take the position that quantum states are real, but reductions are artifacts that arise from treating open quantum systems in isolation without explicitly modeling effects due to interactions with the environment. Just as a strict subjectivist would view the mixing parameter that appears in de Finetti's theorem as a convenient modeling fiction, some scientists prefer to treat reductions as modeling fictions. There have been many attempts at formulating quantum theory in a manner that does away with reductions, but no such attempt has gained universal acceptance, and there is strong disagreement over whether the objective is achievable.

One such ontology, known as *many worlds*, asserts that a quantum system actually realizes *all* possibilities open to it, but each occurs in a separate reality inaccessible to the other realities. There is a copy of each of us in all the different realities, but only the copy in this particular reality has the experiences associated with this reality. The many worlds interpretation is common in the field of quantum computing. According to another ontology known as *consistent histories* (Omnès, 1999), quantum theory defines a probability distribution on *histories* of a quantum system. A history consists of a product of time-ordered projection operators. A set of histories is *consistent* if the elements of the set are mutually orthogonal, i.e., distinguishable from one another. A family of consistent histories represents a set of alternative possibilities for the trajectory of the system. Given any family of consistent histories, quantum theory defines a unique probability distribution on histories in the family. The consistent histories interpretation is said to do away with the postulate of state reduction. However, it is necessary to augment the usual rules of quantum theory with some means of specifying the family of consistent histories on which probabilities are to be defined. The consistent histories ontology is silent about how this should be done. The selection rule for histories plays the same role in the consistent histories interpretation as the rule for selecting a sequence of reduction operators does in the standard interpretation. Yet another interpretation is the *pilot wave* ontology (Bohm and Hiley, 1993), a nonlocal deterministic theory that includes both classical-like particles and a wave function that guides

their evolution. Like many worlds and consistent histories, the pilot wave ontology regards reduction events as artifacts of incomplete knowledge of observers. Unlike many-worlds and consistent histories, in which the state evolves by unitary evolution alone, the pilot wave interpretation adds to the standard account a description of how the particle's evolution is guided by the pilot wave.

Taking a literal interpretation, the random variable template *ReductionOp*(*sys*, *t*) of Figure 10 represents the reduction operator applied to a given quantum system at a given time. A scientist who views reductions as modeling fictions would agree that the statistical assertions encoded in the MFrags of Figure 10 faithfully reflect the experience of observers. Such a scientist would argue, however, that the random variable template *ReductionOp*(*sys*, *t*) is a modeling fiction that generates accurate phenomenological predictions, but does not correspond to any fundamental physical process.

Other scientists take a realist view of reductions. According to the realist view, the influence of *ReductionOp*(*sys*, *t*) on *State*(*sys*, *t*) and *Measurement*(*sys*, *t*) in Figure 10 is causal. That is, intervening to set the value of *ReductionOp*(*sys*, *t*) to *None* causes *State*(*sys*, *t*) to remain unchanged at the value *PreState*(*sys*, *t*) and *Measurement*(*sys*, *t*) to have a null value indicating a nonexistent measurement. Intervening to set the value of *ReductionOp*(*sys*, *t*) to a particular reduction operator (denoted in Figure 10 by the variable *rdc*) causes one of the possible outcomes of *Measurement*(*sys*, *t*) to occur and *State*(*sys*, *t*) to become the projection of *PreState*(*sys*, *t*) onto the subspace associated with the observed value of *Measurement*(*sys*, *t*). The probabilities associated with a given outcome and post-reduction state are determined by the fiducial probabilities for *PreState*(*sys*, *t*) and the subspace corresponding to the outcome. That is, intervening to apply a particular reduction operator causes different propensities for values of *State*(*sys*, *t*) than if a different reduction operator were applied, or if there were no reduction.

Some scientists theorize that at least some reduction operators occur because of deliberate actions taken by agents. Penrose (e.g., 1994) hypothesizes that reduction represents the singling out of an actual event to occur by a mechanism related both to consciousness and gravitation. Stapp (1999 and Schwartz et al., 2005) also takes a realist view of reduction, but does not implicate gravity as a causal factor. Stapp's ontology is closely related to the measurement theory first proposed by von Neumann (1932) and further developed by Wigner (1967). This ontology hypothesizes that the universe contains systems that can cause state reductions. These *reducing agents* can choose, within as yet to be determined physical limits, when to initiate reductions and which kinds of reductions to initiate. One reasonable hypothesis is that a reducing agent can apply reductions that operate on a subset of the degrees of freedom of its own state (e.g., the degrees of freedom corresponding to the controller for its motor subsystem). The manner in which a reducing agent selects a reduction operator might depend on the state of the reducing agent, which might in turn depend on the outcomes of previous reductions. This provides a means, consistent with the laws of physics, for reducing agents to make efficacious
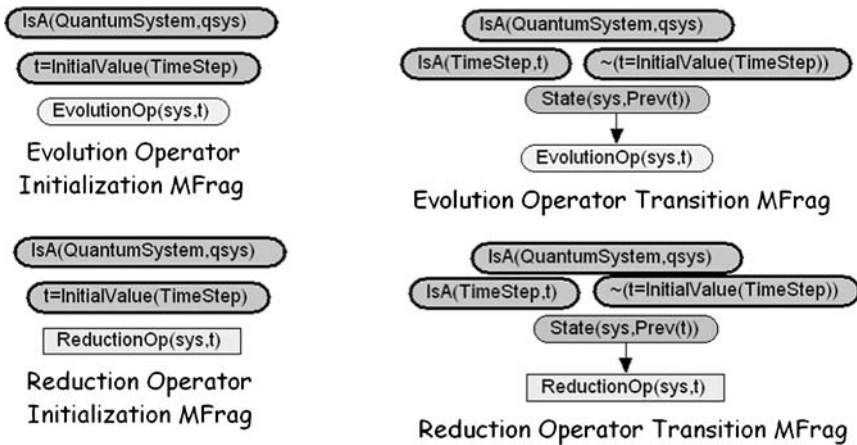
Figure 12. MFrags for operator application by reducing agents.

choices that depend on memories encoded in their quantum states. Because this choice process conforms to the precepts of quantum theory, the choices of reducing agents are consistent both with the constraints imposed by relativity theory that preclude faster than light influences in unitary evolution, and with the nonlocal effects quantum theory predicts for outcomes of reductions. Figure 12 shows an alternate version of the operator generation MFrags of Figure 11, in which reduction operators are modeled as decision nodes.

Like the consistent histories ontology, the reducing agent ontology specifies a probability distribution for a family of consistent histories. The actual history is chosen from this family according to the usual rule (described in Section 3.1 above) for assigning probabilities to outcomes of reductions. It is important to note that in the reducing agent ontology, the choice and timing of reductions may depend on the outcome of previous reductions, if these outcomes are recorded in the memory of the reducing agent.

Stapp (1999) Schwartz et al. (2005) hypothesizes that human agents are one kind of reducing agent. Under this hypothesis, choices by human agents give rise to interventions that cause a quantum system to behave differently from how it would have behaved without the interventions. The hypothesis that humans are reducing agents fills complementary explanatory gaps in physics and psychology by postulating an interaction between the informational structure represented by the quantum state and the informational structure of conscious experience (Stapp, personal communication). Thus, with no change to the mathematical machinery of quantum theory, the reducing agent ontology connects physical reality in a plausible way to conscious experience and goal-directed, deliberate choice.

The reducing agent ontology has empirical consequences regarding the manner in which volition operates in a quantum world. In particular, it is postulated that agents act on the world by initiating reductions applied to *their own* states, and

affect other systems only indirectly through the effects of these interventions. According to the reducing agent ontology, volition operates on the material world via the application of reduction operators. In particular, the reducing agent ontology postulates that the volitional aspects of quantum computing (i.e., temporal evolution of unitary operators programmed as steps in a quantum algorithm; preparation of initial states in a given configuration) must be brought about through the effects of applying reduction operators.

Stapp (1999) and McFadden (2000) describe empirically verified examples of macroscopically detectable differences in behavior resulting from different policies for effecting state reductions in quantum systems. The quantum Zeno effect (Itano et al., 1990; Gribbin, 1996) predicts that observations taken sufficiently rapidly can keep a quantum system within a constrained region of state space. The inverse quantum Zeno effect (McFadden, 2000) induces a quantum system, via a sequence of rapidly repeated measurements, to follow a particular path in its state space. Stapp argues that an organism might use the quantum Zeno effect to keep its brain state within a given basin of attraction sufficiently long to trigger behaviors the organism desires to bring about. The quantum Zeno effect has been confirmed experimentally (Itano et al., 1990) and is thought to occur at time and frequency scales consistent with patterns of electrochemical activity occurring in brains.

These results demonstrate that applying different reduction policies can result in macroscopically distinguishable differences in behavior. This suggests the possibility of empirical tests that could distinguish the reducing agent ontology from ontologies that treat reductions as artifacts or from a realist ontology in which the timing and choice of reductions do not depend on the state of the system prior to reduction.

Although both Stapp and McFadden hypothesize that humans are reducing agents, there is no implication that humans are the only reducing agents or that reducing agents must be conscious. Reductions that occurred prior to the evolution of conscious organisms would have been caused by unconscious or proto-conscious reducing agents. Although it is conceivable that some form of the property we call consciousness at the human level exists throughout the natural world, the reducing agent ontology does not require it.

Treating reductions as real processes that can be initiated by conscious agents provides a theory for aspects of quantum computing that are now treated purely phenomenologically. For example, consider the problem of setting the input state to a quantum algorithm to a given value. As noted above, the ability to specify initial states is taken as a given in the theory of quantum computing. Nevertheless, preparation of initial states is considered to be a difficult issue in the design of practical quantum computing devices (Nielsen and Chuang, 2000). The quantum Zeno effect provides a mechanism, consistent with the reducing agent ontology, for preparing a desired quantum state. Suppose, for example, that we wish to prepare the system in a state in which all qubits have value $|0>$. A procedure to prepare this state can be defined as follows:

Procedure Z (Qubit Zero State Preparation):

1. Measure all qubits in their single-qubit fiducial basis (i.e., measure whether each qubit is in the state |0> or |1>).
2. For each qubit:

   a. If the result of the previous measurement was |0>, measure the single-qubit fiducial basis again.
   b. If the result was |1>, measure one of the fiducial non-basis states (this will yield |0> or |1> with equal probabilities).

3. If all qubits have value |0>, output the result. Else, go to Step 2.

Applying this algorithm will eventually produce the desired state with probability 1, provided that reductions are applied sufficiently rapidly with respect to the rate of change of the state under unitary evolution. A similar kind of algorithm can be applied, given the kind of conditions described in Section 3.2.6 above, to design systems that behave as if a given unitary operator has been 'applied' to the system.

Although Procedure Z works in theory, it remains to be determined whether it could be turned into a practical method for preparing initial states. An important engineering issue is controlling the effects of environmental decoherence.

4. Quantum Agents

It has been argued (e.g., Pearl, 1988; Russell and Norvig, 2002; Lee and Mumford, 2003; Binford and Levitt, 2003) that intelligence requires the ability to perform the functional equivalent of approximate Bayesian reasoning. Graphical probability and decision models are attractive as a logically consistent language for formulating theories of computational intelligence and developing computer implementations capable of approximately optimal inference and decision-making. The Bayesian logic presented in Section 2 is capable of representing a joint probability distribution over truth-values of virtually any set of scientific hypotheses. Given a MEBN theory, Bayesian inference can be applied to compute responses to probabilistic queries, to incorporate empirical evidence and revise beliefs, and to learn improved representations from observations. MEBN logic is capable of representing a probability distribution over models of any first-order theory, and as shown in Section 3, can represent theories of quantum systems and their interaction with systems described classically.

While current implementations of decision theoretic agents run on digital computers, future implementations of decision theoretic agents might employ quantum hardware. One might ask why quantum theory should be important to a theory of physical symbol systems. Present theories of computational intelligence, with their basis in digital computers, have taken us a long way in our twin quests to understand

intelligence and to develop intelligent systems. Additional advances are occurring at a rapid rate. Many prominent neuroscientists are skeptical of claims regarding the relevance of quantum theoretic effects to the study of human and animal cognition. So why, one might ask, should the field of computational intelligence pay attention to quantum theory?

One response to this question is to note that while graphical models have brought computational probabilistic reasoning within the range of feasibility, many problems of interest are beyond the capability of the most efficient known algorithms. The most effective algorithms for highly complex problems are stochastic (e.g., Cheng and Druzdzel, 2000; Doucet et al., 2001). Most current implementations of stochastic algorithms employ pseudo-random numbers. It is well known (e.g., Marsaglia, 1968) that pseudo-random numbers can fail to be random, causing serious but difficult to detect inaccuracies in the results of computations that depend on randomization. Physical randomness has been employed to counteract this problem. Quantum devices enable realizations of randomized algorithms that are physically accurate to the limits of current physical knowledge. Of course, randomized algorithms, even those employing quantum randomization devices, are not the same as true quantum computing. It has been shown that quantum computers can solve important classes of problems with far less computational resources than the best-known classical algorithms (e.g., Aharanov, 1999; Grover, 1996; Shor, 1994), and can perfectly simulate arbitrary finite-dimensional physical systems (Deutsch, 1985). Quantum computers are believed to be intrinsically more powerful than classical algorithms with randomization. Spector suggests that efficient inference in Bayesian networks is a promising application area for quantum computing research (Spector et al., 1999). It is conceivable that quantum algorithms for inference and optimization in graphical probability and decision models might improve on the performance of the best current methods.

More fundamentally, the view of intelligence as algorithms running on digital computers, while it has achieved pragmatic success as a working approximation, is fundamentally inadequate as a foundational theory of computational intelligence. Boolean logic has proven unsatisfactory as a foundational logic for intelligent systems, and is being superceded by Bayesian logic. First-order Bayesian logic is sufficiently powerful to represent virtually any scientific hypothesis, has a theoretically principled means of refining theories based on observation, and provides a principled basis for decision making under uncertainty. Furthermore, we saw in Section 3 above that first-order Bayesian logic can represent the evolution of quantum systems, whereas the family of computable functions is insufficiently rich to represent many important types of physical system. In other words, a Bayesian logic implemented in a quantum device is in principle capable of learning a faithful representation of itself and the quantum world it inhabits. This cannot be said of classical logic implemented on a digital computer.

The reducing agent ontology, coupled with recent innovations in probabilistic knowledge representation, suggests an intriguing perspective on von Neuman's

(1932) seminal formulation of quantum measurement. Von Neumann's formal theory of measurement embeds observer and observed system in a larger system whose evolution is governed by unitary transformations punctuated by reduction events. Von Neumann believed that reduction could not be eliminated by treating it as an artifact of incomplete knowledge:

> . . . the measurement or the related process of the subjective perception is a new entity relative to the physical environment and is not reducible to the latter. (1932, p. 418)

The passage continues with a definition of the principle of psycho-physical parallelism, which von Neumann saw as fundamental to a scientific treatment of the process by which scientists gain knowledge about the world:

> Indeed, subjective perception leads us into the intellectual inner life of the individual, which is extra-observational by its very nature (since it must be taken for granted by any conceivable observation or experiment). Nevertheless, it is a fundamental requirement of the scientific viewpoint – the so-called principle of the psycho-physical parallelism – that it must be possible so to describe the extra-physical process of the subjective perception as if it were in reality in the physical world – i.e., to assign to its parts equivalent physical processes in the objective environment, in ordinary space. (1932, p. 418–9)

To assign physical referents to the subjective components of inner experience, it is necessary to clearly delineate the correspondence between the measured phenomenon and the experience of the observer. Von Neumann noted that the boundary between observer and observed is to a large extent arbitrary, but different ways of delineating the boundary must be mutually consistent. One of von Neumann's most important contributions was his proof that observable outcomes do not depend on the manner by which a composite quantum system is divided into "observed" and "observer" subsystems. That is, predictions are identical whether we localize the reduction event in the brain of the observer, in the sensory organs of the observer, or at the measuring instrument itself. Thus, debates over whether the reduction event "really" occurs in the observer's brain or at the measuring instrument cannot be settled by empirical data.

 Nevertheless, von Neumann was firm in his conviction that we must put the boundary *somewhere*. That is, we cannot do without reductions altogether, if the physical referents of the theory are to have any connection with the experience of observers.

> . . . no matter how far we calculate – to the mercury vessel, to the scale of the thermometer, to the retina, or into the brain . . . at some point, we must say: and this is perceived by the observer. . . . in each method of description, the boundary

must be put somewhere, if the method is not to proceed vacuously, i.e., if a comparison with experiment is to be possible. Indeed experience only makes statements of this type: an observer has made a certain (subjective) observation; and never any like this: a physical quantity has a certain value. (1932, p. 420)

To make statements about the physical world, von Neumann argued, we must correlate the inner subjective experiences of observers with elements of the theory that correspond to the referents of our statements. To connect the experience: "I saw the dial pointing at the number seven," with the physical hypothesis that an object has a certain velocity, requires a mathematical model of the connection between the observer's subjective experience of "seeing seven," the physical state of the observer's cognitive system when the observer "sees seven," the events in the observer's sensory apparatus that occur when the dial is pointing at seven, the position of the pointer on the measuring device, and the physical state of the system whose velocity is being measured.

The science of von Neumann's time had no formal mathematical theories governing "the inner intellectual life of the individual." The intervening years have seen great strides in cognitive psychology, neuroscience, and artificial intelligence. We are still very far from any direct understanding of the relationship between brain states and the contents of consciousness. On the other hand, current technology has produced robots that, at least in restricted domains, can operate successfully in reasonably complex environments. They can process sensory inputs, reason at the cognitive level about goals and plans, combine cognitive-level and sensory-level reasoning to predict the effects of actions, navigate in their environment to carry out their plans, and learn from experience. We have achieved a good understanding of the relationship a robot's internal state bears to the cognitive-level description employed by the software engineers who design its algorithms. Thus, in robotics at least, we can carry through the connection from the state of the robot's cognitive system when "seeing seven" to sensory inputs when "seeing seven" to the pointer on the dial. Obviously, we have no way to know whether it is meaningful to speak of the robot's inner subjective experience when its cognitive system is in a state of "seeing seven." Nevertheless, we can program a robot to behave as if it had a certain set of goals, and to act in an appropriate goal-directed way when it is in a state of "seeing seven."

All this is well and good, but today's robots employ digital computers. So why are we placing so much emphasis on quantum computers?

First, if quantum theory is correct, today's robots do employ quantum computers, because all physical devices are quantum devices. Second, while today's robots are far more primitive than Lt. Cmdr. Data, the fictional robot pictured in Figure 1, many scientists subscribe to the view that quantum computing will be an essential technology for intelligent systems of the future. If systems of this level of sophistication can be achieved at all, many scientists believe intrinsically quantum sensory and computing systems will be a necessary ingredient in their design. Although this

viewpoint is by no means universal, neither can it be dismissed out of hand. It is reasonable to hypothesize that the internal cognitive architecture of a robot of the future might consist of a quantum hardware instantiation of a first-order Bayesian logic. A decision support robot of the future might employ an extended and more fully-developed version of a MEBN theory such as Figure 4. This robot might apply a more complex version of the situation-specific model of Figure 5 to reason about individual situations in its domain of responsibility.

Such a robot might even reason about intrinsically quantum phenomena. As a concrete example relevant to our case study, Suliban cloaking devices are said to employ particle radiation to alter the molecular structure of matter, and can be detected by a device called a quantum beacon.[7] We might add a set of MFrags to our Starship theory to model the use of quantum beacon sensors to detect cloaked Suliban starships. These MFrags would explicitly represent quantum effects of the interaction between Suliban cloaking devices and quantum beacon detectors. The MFrags of Section 3 above might be employed in the robot's internal representation of quantum beacon detectors.

Suppose Lt. Cmdr Data were to prepare a quantum beacon detector, take an observation, and report to Captain Picard that he had observed a sensor reading indicating the presence of cloaked starships in the vicinity. Captain Picard would have the subjective experience of hearing Data give his report. Tracing along von Neumann's moveable cut, we could connect this experience of a verbal report to the sound waves that emanated from Data's vocal system. From there, we could follow the chain to the internal state of Data's cognitive system as he made the decision to vocalize his report. From there we could follow the chain to the state of his sensing apparatus, and finally to the state of the quantum beacon readout. Individuals might differ about whether, in such a scenario, it would be reasonable to ascribe to Data the subjective experience of seeing the sensor reading. Whatever one's view on the truth of the matter, the crew's survival depends on the degree to which Data behaves "as if" he is having the experience in question. Data's designers will build him to act as if he has the goal of providing the Captain with the information he needs to respond appropriately when the ship is in danger.

Regardless of one's metaphysical stance on the objective reality of reductions, the MEBN theory presented here can be used to define a fully Bayesian model for reasoning about quantum systems, and for revising a model of a given quantum system in the light of observations generated by the system. The quantum theory MFrags can also be combined with MFrags that do not explicitly incorporate quantum-level effects. Such a combined model can be used to reason about systems for which some aspects can be treated classically, while others require explicit quantum-level modeling. In principle, we could explicitly model the quantum aspects of other systems represented by our theory (e.g., we could represent

---

[7] See http://www.memory-alpha.org/en/wiki/Cloaking_device for more information on cloaking technology and devices for detecting cloaked starships.

distance sensors quantum-mechanically). However, this would involve prohibitive computational overhead, for no discernable gain in accuracy.

In summary, a formal theory of Bayesian logic running on quantum hardware provides a sound basis for a science of physically embodied cognitive agents. The correspondence between the quantum level description of the physical state of the agent's cognitive system and the cognitive level description of the agent's knowledge provides a formal mathematical model of von Neumann's principle of psycho-physical parallelism. Of course, no mathematical model can actually *be* an experience. Nevertheless, the correspondence between the cognitive and quantum level descriptions is a formal mathematical model of the correspondence between the inner subjective experience of an ideal observer and the physical state of its cognitive apparatus. The theory described here thus extends von Neumann's "movable cut" further toward the cognitive realm than has heretofore been possible. The foregoing arguments are not intended to imply a literal interpretation of human brains as quantum Bayesian computers, any more than the now-standard approach of implementing cognitive theories on digital computers implies a literal belief that human brains are digital computers. Nevertheless, the cognitive architecture described here has the potential to provide new insights into human cognition, artificial intelligence, quantum computing and the relationship between the symbolic and sub-symbolic levels of description.

Acknowledgments

References

Aharanov, D., 1999, "Quantum computation," in *Annual Reviews of Computational Physics VI*. D. Stauffer (ed.), Singapore: World Scientific.

Albert, D., 1992, *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press.

Anderson, J.R., 1999, *Cognitive Psychology and Its Implications*. New York: Worth Publishing.

Bangsø, O. and Wuillemin, P.H., 2000, *Object Oriented Bayesian Networks: A Framework for Top-down Specification of Large Bayesian Networks and Repetitive Structures*. Technical Report CIT-87.2-00-obphw1. Aalborg: Department of Computer Science, Aalborg University.

Berg, J., Nelson, F., and Rietz, T., 2001, *Accuracy and Forecast Standard Error of Prediction Markets*. University of Iowa, College of Business Administration.

Binford, T. and Levitt, T.S., 2003, "Evidential reasoning for object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(7), 837–851.

Blume, L., Easley, D., and Halpern, J.Y., 2005, *Decision Theory Without 'Acts'*. Invited Presentation. Edinburgh, UK: Twenty-First Conference on Uncertainty in Artificial Intelligence. http://www.santafe.edu/~leb/EdinburghSlides.pdf.

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

Bohm, D. and Hiley, B., 1993, *The Undivided Universe*. London, UK: Rutledge.

Bohr, N., 1934, *Atomic Theory and the Description of Nature*. Cambridge, UK: Cambridge University Press.

Boutilier, C., Dean, T., and Hanks, S., 1999, "Decision-theoretic planning: Structural assumptions and computational leverage," *Journal of Artificial Intelligence Research*, **11**, 1–94.

Buntine, W.L., 1994, "Operations for learning with graphical models," *Journal of Artificial Intelligence Research* **2**, 159–225.

Cheng, J. and Druzdzel, M., 2000, "AIS: An adaptive importance sampling algorithm for evidential reasoning in belief networks," *Journal of Artificial Intelligence Research*, **13**.

Costa, P., 2005, Bayesian Semantics for the Semantic Web. Doctoral Dissertation, Fairfax, VA: School of Information Technology and Engineering. George Mason University. http://www.pr-owl.org/bayesianweb.

Dawid, A.P. and Vovk, V.G., 1999, "Prequential probability: Principles and properties," *Bernoulli*, **5**, 125–162.

de Finetti, B., 1974–75, *Theory of Probability: A Critical Introductory Treatment*. New York: Wiley.

DeGroot, M.H., 1970, *Optimal Statistical Decisions*. New York: McGraw Hill.

Deutsch, D., 1985, "Quantum theory, the Church-Turing principle and the universal quantum computer," *Proceedings of the The Royal Society of London A*, **400**(1818), 97–117.

Doucet, A., de Frietas, N., Gordon, N., and Smith, A. (eds). 2001, *Sequential Monte Carlo Methods in Practice*. Berlin: Springer-Verlag.

Druzdzel, M.J. and Simon, H.A., 1993, "Causality in Bayesian belief networks," *Uncertainty in Artificial Intelligence: Proceedings of the Ninth Conference*, San Francisco, CA, Morgan Kaufmann.

Elliott, R.J., Aggoun, L., and Moore, J.B., 1995, *Hidden Markov Models: Estimation and Control*. Berlin: Springer-Verlag.

Fienberg, S.E. and DeGroot, M.H., 1982, "Assessing probability assessors: Calibration and refinement," pp. 291–314 in *Statistical Decision Theory and Related Topics III*. S.S. Gupta and J.O. Berger (eds.), Academic Press.

Fine, T.L., 1973, *Theories of Probability*. New York, NY: Academic Press.

Getoor, L., Friedman, N., Koller, D., and Pfeffer, A., 2001, "Learning probabilistic relational models," in *Relational Data Mining*. Saso Dzeroski and Nada Lavrac (eds.), Berlin: Springer-Verlag.

Gigerenzer, G., Todd, P., and Group, T.A. (eds). 1999, *Simple Heuristics that Make us Smart*. Oxford, England: Oxford University Press.

Gilks, W., Thomas, A., and Spiegelhalter, D.J., 1994, "A language and program for complex Bayesian modeling," *The Statistician* **43**, 169–178.

Grenander, U., 1996, *Elements of Pattern Theory*. Baltimore, MD: Johns Hopkins University Press.

Gribbin, J., 1996, *Schrödinger's Kittens and the Search for Reality*. London: Phoenix.

Grover, L., 1996, "A fast quantum mechanical algorithm for database search," *28th Annual ACM Symposium on the Theory of Computation*, New York, ACM Press.

Hardy, L., 2001, *Quantum Theory from Five Reasonable Axioms*. quant-ph/0101012.

Heckerman, D., Geiger, D., and Chickering, D.M., 1995, "Learning bayesian networks: The combination of knowledge and statistical data," *Machine Learning* (20), 197–243.

Heisenberg, W., 1958, "The representation of nature in contemporary physics," *Daedalus* **87**, 95–108.

Howson, C. and Urbach, P., 1993, *Scientific Reasoning: The Bayesian Approach*. Chicago, IL: Open Court.

Itano, W., Heinzen, D., Bollinger, J., and Wineland, D., 1990, "Quantum zeno effect," *Physical Review A* **41**(5), 2295–2300.

Jensen, F.V., 2001, *Bayesian Networks and Decision Graphs*. Berlin: Springer-Verlag.

Kearns, M. and Mansour, Y., 2002, "Efficient nash computation in large population games with bounded influence," *Uncertainty in Artificial Intelligence Conference: Proceedings of the Eighteenth Conference*, San Mateo, CA, Morgan Kaufmann Publishers.

Koller, D. and Pfeffer, A., 1997, "Object-oriented bayesian networks," *Uncertainty in Artificial Intelligence: Proceedings of the Thirteenth Conference*, San Francisco, CA, Morgan Kaufmann.

Korb, K.B. and Nicholson, A.E., 2003, *Bayesian Artificial Intelligence*: Chapman and Hall.

Langseth, H. and Nielsen, T., 2003, "Fusion of domain knowledge with data for structured learning in object-oriented domains," *Journal of Machine Learning Research* **4**, 339–68.

Laskey, K.B., 2005, "First-order bayesian logic," Fairfax, VA: Department of Systems Engineering and Operations Research, George Mason University.

Laskey, K.B. and Costa, P., 2005, "Of klingons and starships: Bayesian logic for the 23rd century," *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-first Conference*, Arlington, VA, AUAI Press.

Lee, T.S. and Mumford, D., 2003, "Hierarchical Bayesian inference in the visual cortex," *Journal of the Optical Society of America A*, **20**(7), 1434–48.

Levitt, T.S., Winter, C.L., Turner, C.J., Chestek, R.A., Ettinger, G.J., and Sayre, S.M., 1995, "Bayesian inference-based fusion of radar imagery, military forces and tactical terrain models in the image exploitation system/balanced technology initiative," *International Journal of Human-Computer Studies* **42**.

Lewis, D., 1980, "A subjectivist's guide to objective chance," pp. 263–293 in *Studies in Inductive Logic and Probability*. R.C. Jeffrey (ed.), Berkeley, CA: University of California Press.

Mahoney, S.M. and Laskey, K.B., 1998, "Constructing situation specific networks," *Uncertainty in Artificial Intelligence: Proceedings of the Fourteenth Conference*, San Mateo, CA, Morgan Kaufmann.

Marsaglia, G., 1968, "Random numbers fall mainly in the planes," *Proceedings of the National Academy of Science*, **61**, 25–28.

Martignon, L. and Laskey, K.B., 1999, "Taming wilder demons: Bayesian benchmarks for fast and frugal heuristics," in *Simple Heuristics that Make us Smart*. Gerd Gigerenzer, Peter Todd and The ABC Group (eds.), Oxford University Press.

McFadden, J.J., 2000, *Quantum Evolution*. New York, NY: Norton.

Murphy, K. and Russell, S., 2000, "Rao-blackwellised particle filtering for dynamic bayesian networks," *Uncertainty in Artificial Intelligence: Proceedings of the Sixteenth Conference*, San Mateo, CA, Morgan Kaufmann.

Nau, R.F. and McCardle, K.F., 1991, "Arbitrage, rationality, and equilibrium," *Theory and Decision*, **31**, 199–240.

Neapolitan, R.E., 2003, *Learning Bayesian Networks*. New York: Prentice Hall.

Newell, A. and Simon, H., 1976, "Computer science as empirical inquiry: Symbols and search," *Communications of the ACM* **19**(3).

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

Omnès, R., 1999, *Understanding Quantum Mechanics*. Princeton, NJ: Princeton University Press.

Parker, R.C. and Miller, R.A., 1987, "Using causal knowledge to create simulated patient cases: The CPCS project as an extension of Internist-1," *Computer Applications in Medical Care: Proceedings of the Eleventh Annual Symposium*, IEEE Comp Soc Press.

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

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

Pfeffer, A., 2000, *Probabilistic Reasoning for Complex Systems*. Stanford, CA. Stanford University

Poole, D., 2003, "First-order probabilistic inference," *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*.

Pratt, J.W., Raiffa, H., and Schlaifer, R., 1965, *The Foundations of Decision Under Uncertainty: An Elementary Exposition*. New York: McGraw Hill.

Russell, S. and Norvig, P., 2002, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice-Hall.

Savage, L.J., 1954, *The Foundations of Statistics*. New York: Wiley.

Schwartz, J.M., Stapp, H.P., and Beauregard, M., 2005, "Quantum physics in neuroscience and psychology: A new model with respect to mind/brain interaction," *Philosophical Transactions of the Royal Society B*, **360**(1458), 1309–27.

Shafer, G., 1981, "Constructive probability," *Synthese* **48**, 1–60.

Shafer, G. and Vovk, V., 2001, *Probability and Finance: It's Only a Game*. New York: Wiley.

Shor, P.W., 1994, "Algorithms for quantum computation: Discrete logarithms and factoring," *35th Annual Symposium on the Foundations of Computer Science*, Los Alamitos, CA, IEEE Press.

Spector, L., Barnum, H., Bernstein, H.J., and Swamy, N., 1999, "Quantum computing and artificial intelligence: Abstract for invited presentation," *Proceedings of the Sixteenth National Conference on Artificial Intelligence, AAAI-99*, AAAI Press.

Spiegelhalter, D.J., Thomas, A., and Best, N., 1996, "Computation on graphical models," *Bayesian Statistics*, **5**, 407–25.

Stapp, H., 1999, "Science of consciousness and the hard problem," *The Journal of Mind and Behavior* **18**, 171–93.

Takikawa, M., d'Ambrosio, B., and Wright, E., 2001, "Real-time inference with large-scale temporal Bayes nets," *Uncertainty in Artificial Intelligence: Proceedings of the Seventeenth Conference*, San Mateo, CA, Morgan Kaufmann.

Tarski, A., 1944, "The semantic concept of truth and the foundations of semantics," *Philosophy and Phenomenological Research*, **4**.

Thagard, P., 1988, *Computational Philosophy of Science*. Cambridge, MA: MIT Press.

von Neumann, J., 1932, *Mathematical Foundations of Quantum Mechanics* (Chapter VI), Princeton, NJ: Princeton University Press.

von Winterfeldt, D. and Edwards, W., 1986, "Decision Analysis and Behavioral Research."

Wigner, E., 1967, "Remarks on the mind-body problem," *Symmetries and Reflections* 171–184.

Williamson, J., 2004, "Philosophies of probability: Objective Bayesianism and its challenges," in *Handbook of the Philosophy of Mathematics*, Dordrecht, Netherlands: Elsevier.