AFRL-AFOSR-VA-TR-2019-0247

Applications of Quantum Probability Theory to Strategic Decision Making

Jerome Busemeyer
TRUSTEES OF INDIANA UNIVERSITY

08/16/2019
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
19-08-2019

2. REPORT TYPE
Final Performance

3. DATES COVERED *(From - To)*
30 Sep 2015 to 29 Sep 2019

4. TITLE AND SUBTITLE
Applications of Quantum Probability Theory to Strategic Decision Making

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-15-1-0343

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Jerome Busemeyer

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
TRUSTEES OF INDIANA UNIVERSITY
205C BRYAN HALL 105 N INIANA AVE
BLOOMINGTON, IN 47405-1106 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2019-0247

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The broad and long term goal of this research program is to provide a new foundation for constructing probabilistic-dynamic systems from principles based on quantum as opposed to classical probability theory. To be clear, we are not interested in physics, and neither do we claim the brain is a quantum computer. Our interest lies solely in the application of mathematical principles from quantum theory to cognitive and decision sciences. Our recent research demonstrates that quantum theory provides a viable new direction for organizing and accounting for paradoxical findings from decision research using a unified and principled theoretical framework (e.g., Busemeyer & Bruza, 2012; Busemeyer & Wang, 2014;
Khrennikov, 2010; Wang, Busemeyer, Atmanspacher, & Pothos, 2013).
So far we have applied these principles to both traditional, one-stage decision problems studied by decision researchers as well as dynamic Markov decision problems used in computer science and engineering.
The more specific goal of the proposed research was to extend our initial work by
developing new applications of quantum probability applied to dynamic, strategic, multipleagent decision situations.
Specifically:
(1) We Developed new Markov and quantum models for perspective taking by opposing agents in dynamic, strategic games, and derived theoretical predictions that discriminate between the competing two classes of models.
(2) We experimentally tested the proposed quantum model of perspective taking using probability judgments and decisions for self versus opponents in traditional economic strategic games, and reported evidence for the quantum model.
(3) We develop engineering applications of quantum models to the strategic decision task of multiple agents deciding to acquire multiple targets in an uncertain and dynamically changing environment.

15. SUBJECT TERMS
Decision Models, Autonomous systems

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON LAWTON, JAMES
a. REPORT	b. ABSTRACT	c. THIS PAGE			
Unclassified	Unclassified	Unclassified	UU		19b. TELEPHONE NUMBER *(Include area code)* 703-696-5999

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Previous research by investigators

Our previous research had two major aims. One was to develop applications of quantum probability and quantum dynamics to human decision making behavior. A second was to develop engineering applications of quantum theory to dynamic and strategic decision tasks.

1 Applications to human judgment and decision making

First we review our earliest applications of quantum probability to human judgments and decisions under uncertainty. Second we review later applications to dynamic decisions involving evidence monstering and accumulation across time. Third we review our most recent applications to strategic decision making. An important point that we wish to make is the following. In the past, various different kinds of ad hoc heuristics have been used to account for the various puzzling findings reviewed below – a different heuristic model is made up for each phenomenon. Our goal is to use the same basic quantum principles to account for all the various different findings and connect these phenomena together, which have never been connected in past work.

1.1 Probability judgments errors and irrational decision making

One of our early applications was designed to account for well-known research on probability judgment errors [12]. Two of the most important are the conjunction and disjunction fallacies. A conjunction fallacy occurs when a person judges the probability of the conjunction of two events to be more likely than one of the constituent events. An example would be judging the conjunctive event that a man that is over 50 years old (event O) and has a heart attack (event H) to be more likely than the event that a man has a heart attack. But according to the law of total probability $p(H) = p(H \cap O) + p(H \cap \bar{O}) \geq P(H \cap O)$. A disjunction fallacy occurs when a person judges the probability of the disjunction of two events to be less likely than one of the constituent events. An example would be judging the disjunctive event that a man is over 50 or has a heart attack to be less likely than the event that a man is over 50. We [13] developed a simple but general quantum probability (QP) account for these puzzling findings as follows. We define P_H as the projector for the event H, define P_O as the projector for the event O, $P_{\bar{O}}$ as the projector for the event not old ($P_O \cdot P_{\bar{O}} = 0, P_O + P_{\bar{O}} = I$), and define ψ as the state based on a person's beliefs. Then the quantum probability of event H equals $p(H) = \|P_H \cdot \psi\|^2$ and the quantum probability for the sequence of events O and then H equals $\|P_H P_O \cdot \psi\|^2$. However, we can decompose the probability of event H as $\|P_H \cdot \psi\|^2 = \|P_H P_O \cdot \psi + P_H P_{\bar{O}} \cdot \psi\|^2 = \|P_H P_O \cdot \psi\|^2 + \|P_H P_{\bar{O}} \cdot \psi\|^2 + Int$, where *Int* symbolizes the cross product terms. If *Int* is sufficiently negative then we obtain $\|P_H \cdot \psi\|^2 < \|P_H P_O \cdot \psi\|^2$. A similar application can produce the disjunction fallacy (see [13] for details). Our model of the conjunction and disjunction fallacies was developed after the facts were known. But our theory also made new predictions about these fallacies. One prediction, in particular, was based on the implication from our theory that the events producing these fallacies must be incompatible to produce these fallacies, which implies that the order of judgment of events should matter. Therefore, we predicted that these fallacies should be related to question order effects, which was a prediction that was later supported by an empirical tests [14] (However, see [15] for an opposing argument.)

Using the same principles described above, we developed a general model for question order effects: If the events O and H are incompatible, then the probability of the sequence of answers O and then H, which equals $\|P_H P_O \cdot \psi\|^2$, will differ from the probability for the sequence of events H and then O, which equals $\|P_O P_H \cdot \psi\|^2$. More importantly, we derived a general, *a priori,* parameter free, quantitative prediction from this general model, that we call the QQ equality: *QQ* = [*p*(yes to A then no to B) + *p*(no to A and then yes to B)] - [*p*(yes to B and then no to A) + *p*(no to B and then yes to A)] = 0. This prediction about the pattern of order effects provided a strong empirical test of the general model. Our QQ equality prediction was found to be statistically supported across a wide range of 70 national field studies that examined question order effects [16]. (However, this discovery attracted a lot of interest, and after discovering this finding, two other competing non quantum accounts, [17, 18] were proposed.) We have also applied our quantum model to account for the effects produced by the order that information is presented when make inferences about hypotheses [19]. To be more specific, we developed a low dimensional parametric quantum model (similar to the BAE model described later) to account for these order effects. Then we have quantitatively compared the accuracy of the predictions from the quantum model to previous models of order effects [20]. Using the same number of parameters for both models, we found that the quantum produced more accurate predictions than these earlier models of order effects on inference [19].

Another early application was to findings concerning violations of a basic "rational" axiom of decision-making, called the "sure thing" principle [21]. According to the "sure thing" principe, if you prefer action A over B under state of the world X, and you also prefer action A over B under the complementary state of the world not X, then you should prefer action A over B even if the state of the world is unknown. Shafir and Tversky [22] experimentally tested this axiom using the prisoner dilemma (PD) game. Three conditions were used to test the sure thing principle: In an 'unknown' condition, the player acts without knowing the opponent's action; in a 'known defect' condition, the player is informed that the opponent will defect before the player takes action; and in a 'known cooperate' condition, the player is informed that the opponent will cooperate before the player takes action. According to the sure thing principle, if you prefer to defect, regardless of whether you know your opponent will defect or cooperate, then you should prefer to defect even when your opponent's action is unknown. Participants frequently violated the sure thing principle (Shafir & Tversky 1992)—many players defected knowing the opponent defected and knowing the opponent cooperated, but they switched and decided to cooperate when they did not know the opponent's action. This preference reversal by many players caused the proportion of defections for the unknown condition (.63) to fall below the proportions observed under both of the known conditions (.97 knowing the opponent defected, and .84 knowing the opponent cooperated). We later replicated these surprising findings [23]. These results violate a prediction based on the law of total probability: $p(PD) = p(OD) \cdot p(PD|OD) + p(OC) \cdot p(PD|OC)$, where PD is the event that player defects, OD is the event that the opponent is predicted to defect, and OC is the event the the opponent is predicted to cooperate. According to this law, the probability of defection in the unknown condition must fall between the two known conditions. Pothos and Busemeyer (2009) developed a quantum model to account for these results using the same principles as used to account for the probability judgment errors . We define a projector P_{PD} for the player to decide to defect; another projector P_{OD} representing the event that the opponent will defect, and ψ is the initial state of the player. Then the probability that the player defects in the unknown case equals $\|P_{PD} \cdot \psi\|^2 = \|P_{PD}P_{OD} \cdot \psi\|^2 + \|P_{PD}P_{\bar{OD}} \cdot \psi\|^2 + Int$. The Int term can be negative to produce the observed violation of the prediction from total probability.

To be more specific, we developed a low dimensional quantum model to quantitatively account for the interference effects, called the belief-action entanglement (BAE) model. (The basic form of this model was also used by [19].) The BAE model is a 4 dimensional quantum model spanned by 4 tensor product basis vectors $|B_i\rangle \otimes |A_j\rangle, i = \{\text{predict opponent defects, cooperates}\}$, $j = \{\text{player acts to defect, cooperate}\}$. A 4 dimensional unitary matrix U transforms an initially separated state ψ_I to a final entangled state $\psi_F = U \cdot \psi_I$ that coordinates beliefs with actions. The entanglement is a necessary property of the model needed to produce interference. The unitary is constructed from a Hamiltonian matrix H by a matrix exponential function $U = e^{-i \cdot H}$. The Hamiltonian matrix H in the matrix exponential function is computed from parameters determined by the utilities of the payoffs produced by the actions under each world state (opponents action in this case). The projector P_{PD} (which picks out the action to defect) is then applied to the final state ψ_F to produce the probability of action $\|P_{PD} \cdot \psi\|^2$. We [24] compared the quantitative predictions from our BAE quantum model to a traditional decision model originally developed by Tversky and Shafir [22] at the individual level of analysis for large data set with 33 conditions and 100 participants. Both models used the same number of parameters, and the models were compared at an individual level of analysis using a Bayes factor method. The results of the comparison clearly favored the quantum model over traditional decision model (see, [24] for details).

A different application was designed to account for findings that we discovered using a paradigm that was designed for testing interference effects of categorization on a later decisions [25]. On each trial, participants were shown pictures of faces. They were asked to categorize the faces as belonging to either a "good" guy or "bad" guy group, and they were asked to decide whether to take an "attack" or "withdrawal" action. Two critical conditions were used to test interference effects: In the C-then-D condition, participants categorized the face and then made an action decision; in the D-Alone condition, participants only made an action decision (no categorization response was required). The test of interference was based on a prediction based on the law of total probability: define $p(Attack)$ as the probability to attack in the decision alone condition, define $p(G)$ as the probability to categorize as "good guy" and $p(B)$ as the probability to categorize as "bad guy" for the for the C-D condition, and define $p_T(Attack) = p(G) \cdot p(Attack|G) + p(B) \cdot p(Attack|B)$ as the total probability obtained from the C-D condition. An interference effect is defined as the difference $p(Attack) - p_T(Attack)$. In other words, asking about the category interferes with the final probability of taking the action to attack as compared to not asking about the category. We found systematic interference effects across several experiments [25]. For example, in one of the experiments, when the optimal decision was to attack, we found that $p(Attack) = .69 > p_T(Attack) = .60$. We formulated a quantum model using the same principles described for the previous applications: define P_G as the projector for categorizing the face as "good guy", define P_B as the projector for categorizing the face as "bad guy" and define P_A as the projector for deciding to attack. Then the probability to attack in the D - alone condition equals $\|P_A \cdot \psi\|^2 = \|P_A P_G \cdot \psi\|^2 + \|P_A P_B \cdot \psi\|^2 + Int$, and again the *Int* term accounts for the interference effect. To be more specific, we used the same BAE quantum model (used earlier for the prisoner dilemma game) to account for the interference of categorization on decision. We also quantitatively compared a Markov model [26] to our quantum model. The Markov model could not predict the interference effect; nevertheless, it is unclear whether the quantum or Markov could better predict other properties of the choice data. So we used a generalization test to compare the quantitative predictions of the two models: both models were fit to two payoff conditions using the same number of parameters; then these same parameters were used to make new predictions for a new payoff condition. The quantum model provided more accurate predictions for generalization than the Markov model.

Most recently, we have developed a general theory, called Hilbert space multidimensional (HSM) theory [27, 28], to account for the effects of measurement context on judgments. A measurement context refers to a set of empirical variables that are measured on the same occasion; different contexts are formed by measuring different overlapping subsets of variables [29]. A context effect occurs when measurements are affected by the context in which they appear. This can happen when the reaction to the measurement of one variable changes depending on which other variables are being measured in the same context. To give an empirical example in psychology, consider the following study that investigated the effectiveness of persuasive messages. Participants were asked to judge four attributes: how Persuasive, Believable, Informative, and Likable the messages were perceived to be. A context was formed by asking about a pair of attributes regarding the same message (e.g., Is the message likable and persuasive? Answer yes or no separately to each of the two attributes). Each pair of attributes (each context) produced a 2 x 2 joint frequency table (frequencies of yy, yn, ny, nn pairs of answers). Six different contexts were produced by asking for evaluations of six pairs of attributes (e.g., Is the message likable and persuasive? Is the message believable and persuasive?). We tested for context effects by defining a "context free" model: it was defined by four (P,B,I,L) binary random variables that produced a four-way joint distribution. We then tested whether or not it could reproduce (by marginalization) the 6 two-way tables. We found statistically significant deviations from the context free model, which indicated that the interpretation of a message attribute (e.g., persuasive) changed depending on the other attribute with which it was paired (likable versus believable). We built a quantum model using our HSM program (see link **http://mypage.iu.edu/~jbusemey/quantum/HilbertSpaceModelPrograms.htm**) to account for the context effects. We then quantitatively compared the predictions using our HSM model with Bayesian network models that have previously been used for these kinds of applications. Using the same number of model parameters, the results demonstrated that our quantum model predicted more accurately than the Bayes net models that we examined.

1.2 Dynamic decisions based on evidence accumulation

We have extended our theory and experimental tests of interference effects to dynamic decision type problems involving monitoring and accumulating evidence across time to make a decision. In one of our first studies in this line of research, we used a dynamic signal detection type task in which a decision maker decided whether a target is present or absent based on sampling noisy and uncertain information across time (e.g., to decide whether an enemy is located at a position based on watching a poor and fuzzy image over time). Human performance (accuracy, decision time, and confidence) observed with dynamic signal detection tasks has traditionally been modeled using Markov type of random walk or diffusion models of decision-making (e.g., see [30]). The basic idea is that the decision maker accumulates evidence for each hypothesis until the accumulated evidence reaches a threshold. The first hypothesis to reach the threshold is chosen and the time to reach the threshold determines the decision time, and the difference in evidence soon after the decision determines the confidence. Alternatively, we [31] developed a quantum random walk model for signal detection, which assumes that a person's evidence state is represented by a wave function spread over levels of evidence. The Markov model evolves probabilities over time according to the Kolmogorov forward equation, and the quantum model evolves amplitudes over time according to the Schrödinger equation. We [32] (ch. 8) derived a key prediction that provides a critical method to empirically distinguish and test the two theories. The experiment consists of two conditions: In the choice-confidence condition, the person makes a choice (signal present or absent) at time t_1 and then rates confidence at time t_2; in the confidence-alone condition, the person only provides a confidence rating at time t_2. For both conditions, the focus is on the marginal distribution of confidence ratings (pooled across choices for the choice-confidence condition) that are obtained at time t_2. Confidence is defined as the probability that a signal is present on a scale ranging from 0 = the target is not present, to 50 = undecided, to 100 = the target is present. The Markov model obeys the Chapman-Kolmogorov equation, which is a dynamic form of the law of total probability and predicts no difference between the two conditions. The quantum model predicts that an initial decision interferes with a later confidence rating, making the confidence distributions differ between the two conditions. We [10] empirically tested this prediction and obtained strong support for the predicted interference effect: the interference was statistically significant for seven out of nine participants (each participant contributing over 2500 trials). Furthermore, we compared the quantitative predictions of Markov and quantum random walk models using the same number of parameters and using a Bayes factor method at the individual level of analysis. The results of the model comparison favored the quantum model for 7 out of the 9 participants.

In a follow-up study [11], we examined the sequential effects of one continuous confidence judgment on a later judgment. In this experiment, two confidence ratings were made at a pair (t_1, t_2) of time points. The experiment included three main conditions: (1) requests for confidence ratings at times t_1 = 0.50s and t_2= 1.50s; (2) requests for ratings at times t_1 = 1.50s and t_2 = 2.50s, and (3) requests for ratings at times t_1 = 0.50s and t_2 = 2.50s. First of all, this design provided another test for interference effects by comparing the marginal distribution of probability ratings at time t_2 = 1.50s for condition 1 (pooled across ratings made at time t_1 = 0.50s) with the distribution of ratings at time t_1 = 1.50s from condition 2. Once again, the Markov model predicts no difference between conditions at the matching time points, whereas the quantum model predicts an interference effect of the first rating on the second. Only a minority of participants produced significant interference effects in this study. Apparently, the act of making a resolute decision produces greater interference than an irresolute confidence judgment at the first measurement. Secondly, this design also provided a generalization test for quantitatively comparing the predictions computed from the competing models. The parameters from both models were estimated from the confidence ratings distributions obtained from the first two conditions for each individual; then these same parameters were used to predict the confidence rating distribution for each person on the third condition. The Markov model was again based on the Kolmogorov forward equation, and the quantum model was again based on the Schrodinger equation. We estimated two parameters from the joint distribution (pair of ratings at .5s and 1.5s) obtained from condition 1, and the joint distribution (pair of ratings at 1.5s and 2.5s) from condition 2, separately for each participant. Then we used these same two parameters to predict the joint distribution (pair of ratings 0.5s and 2.5s) obtained from condition 3 for each participant. The results of the new experiment indicated that the ratings of the majority of participants were better predicted by the quantum model than the Markov model.

More recently, we have conducted preliminary research on preference evolution across time. The quantum walk predicts oscillation across time, whereas the Markov model used in previous cognitive research predicts no oscillation (instead it predicts growth to an equilibrium state). We have found evidence for oscillation in preference as predicted by the quantum walk process. Furthermore, this oscillation was found to be dampened by previous choice as compared to no choice, an effect also predicted by the quantum walk model but not by the Markov process.

Figure 1: Mean preference across different time measurements following either an initial choice or no choice.

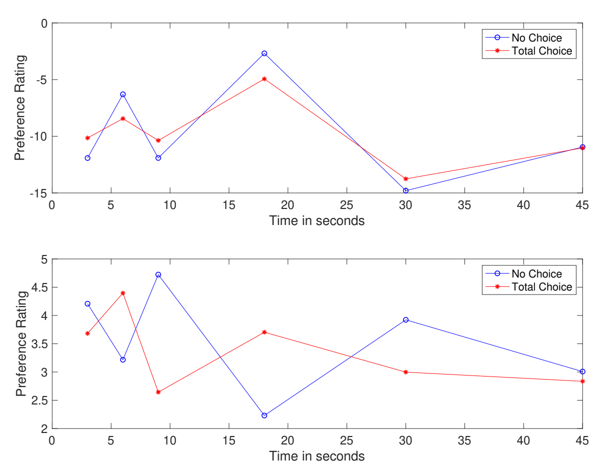


Figure 1 shows the results from two different experiments (top panel is from a consumer preference choice study, and bottom panel is from a study of choice between risky gambles). The plot shows the mean preference plotted across different time measurements after either an initial choice or no choice. The results for the choice condition are pooled across the initial choice to produce a total probability measure of preference. However, even though we have observed oscillation during initial stages, this oscillation eventually disappears. This suggests the need for a hybrid model that includes both quantum and Markov processes. This can be achieved by adding Lindblad operators to the quantum process.

1.3 Applications to strategic decisions

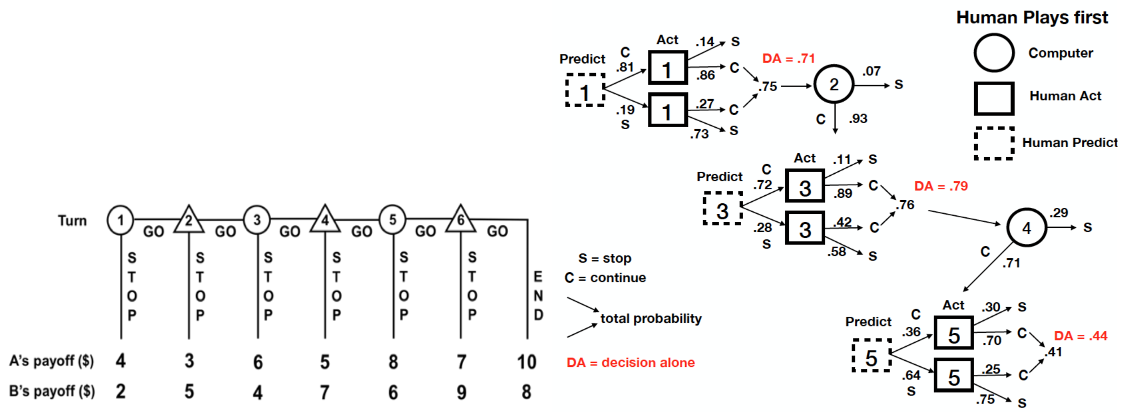
Our most recent work has been to extend our ideas to strategic decision situations situations. In particular, we have designed experiments to test predictions based on our quantum decision models for a sequential game between two players called the centipede game. The centipede game [33] has been extensively studied in the experimental economics literature [34]. The game in extensive form is shown in the left panel of Figure 2: two players, called player A and player B, decide sequentially whether to allocate a pot of rewards in a predetermined way or whether to pass the allocation choice to the other player. If a player passes the decision to the other player, the total pot to allocate increases in size. Passing can be done only a finite number of times (6 in this example). Once a player splits the pot, the game is over with that player gaining the higher share of the pot and the other player obtaining the smaller share. All standard game-theoretic equilibrium concepts predict that the pot is split at the first decision node of the first player (see, e.g., [35]).

The purpose of this new work was to investigate interference effects produced by requesting a prediction (about the opponent's action before making a decision) on the probability of continuation (probability to pass the allocation to the opponent). The basic paradigm is the following: On half the trials, the player first predicts the action of the opponent and then makes a decision (to pass or stop) on each stage; on the other half the trials, the player simply makes a decision on each stage (without any overt prediction). From the predict-act trials, we can compute the total probability of defecting ($p_T(D|PA)$); from the act-alone trials we also observe the probability of defecting without prediction ($p(D|AO)$). The difference between the two $p_T(D|PA) - p(D|AO)$ is the measure of interference, which can be estimated at each of the 6 decision stages. According to standard game theory, the prediction should have no effect on the action, and so there should be no interference. According to our quantum model, as well as past work on the prisoner dilemma game and the categorization - decision task, interference was expected to occur.

In our latest experiment, $N = 493$ participants played the game shown in the left panel of Figure 2 against computer agents programmed to play like humans. Each participant played 30 games with a different computer agent each time (15 with the player going first, and 15 with the agent going first). The player received two types of games: predict-act games on which the player first predicted the action of the agent at each stage and then decided an action; and act-only games on which the agent simply decided an action (without making any prediction).

The right panel of Figure 2 illustrates some of the results when the human player moved first (similar results were obtained when the agent moved first). The figure shows the 5 stages before the end of the game. The proportions shown above the arrows pointing out of the dashed boxes are the predictions by the human regarding whether the agent will continue or stop. The proportions shown above the arrows pointing out of the solid boxes are the actions, following each prediction, by the human player to continue or stop. The proportions after the converging arrows show the total probability to continue at each stage for predict then act games, and the adjacent red text shows the proportion to continue for the decision alone games. As can be seen, interference effects occurred throughout the game (which were statistically significant). The defection rate for the first stage is relatively low in this study, and the interference was negative (more defection with act-only).

Figure 2: Left Panel: Six stage centipede game with linear increasing payoffs. Right Panel: Observed choice probabilities for the condition when the human player moved first.



The defection rate increased on later stages, and the interference effect switched to positive (more defection for predict-act).

The same BAE quantum model that was first developed to quantitatively account for interference effects found with the prisoner dilemma game [36] and the categorization-decision task [25] was extended to apply across the multiple stages of the centipede game. The new model successfully fits the observed data. However, comparisons with traditional game theory models are still in progress (we need to first construct reasonable probabilistic versions of traditional models to compare with the quantum model).

2 Engineering applications

2.1 Quantum reinforcement learning

We have also examined applications of quantum decision theory to more complex predator-prey type of decision problems within the class of Markov decision problems (MDP's) [37, 38]. In particular, we developed a new quantum reinforcement learning algorithm for MDP's. The quantum reinforcement learning algorithm uses the same Q-learning algorithm to estimate values of actions as used in traditional reinforcement learning models [39]. The key difference is concerned with the probabilistic rules to select actions. Unlike traditional models that use, for example, the epsilon greedy algorithm, or the soft max rule for action selection, the quantum model uses quantum probability rules for selecting actions. The basic idea is that the current environmental state puts the agent in a superposition state over the set of possible actions. The key new idea is to use an amplitude amplification algorithm for modifying the superposition state after experiencing a payoff for an action. The amplitude amplification algorithm is an extension of the Grover quantum information search algorithm [40]. This idea was first proposed and tested using simulations by Dong et al. [41, 42]. We have, however, made major modifications to substantially improve Dong's original algorithm. Dong et al. (2008) proposed to map Q values into the parameter L, which is an integer number of amplifications. However, this becomes very problematic for small numbers of actions. Also, this method only amplifies and never attenuates the amplitude assigned to an action. Instead, our new model fixes $L = 1$, and we map normalized values of Q from the Q-learning algorithm into two phases φ_1, φ_2 that moderate the amplification algorithm, which can amplify rewarded actions and attenuate punished actions. The key idea for robustness is that for a given number of actions, m, the mapping from the Q values of the Q-learning model to the parameters φ_1, φ_2 can be determined *a priori* to provide robust learning. Unlike the epsilon greedy and softmax rules, the quantum parameters do not need to be adjusted post hoc for each variation in the environment.

To evaluate our quantum algorithm practically, we conducted computer simulations within a large grid world, using a prey-predator game involving two *competing* predators and one randomly moving prey. The predators were given information about the distance from the prey in each direction on each step. One predator was based on the traditional soft max probabilistic choice rule, and the other was based on our new quantum probabilistic rule. (We also compared results with the epsilon greedy choice rule, but this did not perform as well as the soft max rule.) The aim of the task was to find a policy that will let the predator find the prey with minimum loss. Fakhari et al. [37] conducted extensive simulations varying the size of the grid world and the number of actions. The main results are that at the early stage of training on the task (learning Q values), the soft max algorithm caught more prey than our quantum algorithm; however, at intermediate and later stages, the quantum algorithm strongly outperformed the soft max rule. We have recently been invited to write a chapter summarizing this work in a book edited by Kyriakos Vamvoudakis and Frank Lewis titled "Directions on Reinforcement Learning and Games."

2.2 Stochastic optimal control using path integrals

We developed an offline approximate dynamic programming approach using neural networks for solving a class of finite horizon stochastic optimal control problems [43]. There are two approaches available in literature, one based on stochastic maximum principle (SMP) formalism and the other based on solving the stochastic Hamilton-Jacobi-Bellman (HJB) equation. However, in the presence of noise, the SMP formalism becomes complex and results in having to solve a couple of backward stochastic differential equations. Hence, current solution methodologies typically ignore the noise effect. On the other hand, the inclusion of noise in the HJB framework is very straightforward. Furthermore, the stochastic HJB equation of a control-affine non-linear stochastic system with quadratic control cost function and arbitrary state cost function can be formulated as a path integral problem. However, due to curse of dimensionality, it might not be possible to utilize the path integral formulation for obtaining comprehensive solutions over the entire operating domain. A neural network structure called 'adaptive critic design' paradigm is used to effectively handle this difficulty. We developed a novel adaptive critic approach using the path integral formulation for solving stochastic optimal control problems. The potential of the algorithm was demonstrated through simulation results from a couple of benchmark problems.

2.3 Quantum Inspired Techniques and Multiple-Model Filters

Fault detection and parameter identification are critical issues in the stability of systems with model based control if the model parameters change during the system operation. Detection of faults and adaptive estimation of parameters to be used in control are crucial to successful operation of a vehicle. We developed a new paradigm for fault detection and adaptive parameter estimation [44].

The basic building block is a multiple-model Kalman filter. However, as opposed to many existing studies, the value of parameters used is not fixed and the true values need not be constant. In fact, when a fault occurs, parameters of a system change and can change drastically; in order to cope with such changes, a quantum-boost scheme for multiple-model Kalman filter was developed. At each time step, the a priori estimation and covariance of the Kalman filters are weighted with the conditional probability of the parameter set, given the new measurement. The conditional probability is obtained by Bayesian inference. The conditional probability of the parameter set closest to the correct parameter set typically rises with new information and finally converges to one. In fault detection, it is of vital significance that such probability rise and convergence happen quickly. This can be achieved with an extended Grover's algorithm, originally developed for quantum information processing. In our quantum-boost scheme, the conditional probability is updated twice in each time step, once with Bayesian inference and once with extended Grover's algorithm. The probabilities of the different parameter sets can be considered as the weights for the parameter sets. The weighted estimation can be considered as a superposition state and the estimations of the sub-filters with different parameter sets as basis states. Then the probabilities are the amplitude squared of the basis states. The extended Grover's algorithm is suitable for probability update because the Grover operator is unitary and hence the updated probabilities will sum up to 1. A proof was given to assure that the boost under certain conditions can accelerate the convergence of probabilities. Efficacy of the quantum-boosted multiple-model Kalman filter was tested with two examples. In both examples, the quantum boost scheme was seen to accelerate the rise and convergence of the unknown initial parameter and also the changed parameter. With quantum-boosted multiple-model Kalman filter, fast fault detection can be achieved by monitoring the probabilities of the assumed values of the parameters. Figure 3 shows an example of a problem to detect change in a parameter of a dynamic system. The left panel illustrates standard Bayesian inference and the right panel show the effect of the amplification algorithm. The initial learning period occurs before time point 70, a change in system parameters is learned after time point 70. Learning is much faster with the amplification.

Figure 3: Estimating the parameter of a dynamic system. Left panel shows traditional Bayesian learning, right panel shows speed up by amplification. A change in parameter occurs at time point 70.