A Quantum Information Processing Explanation of Disjunction Effects

Jerome R. Busemeyer (Jerome.Busemeyer@afosr.af.mil**)**
Air Force Office of Scientific Research, 875 N. Randolph Street
Arlington, VA 22203 USA

Mervin R. Matthew (mermatth@indiana.edu**) Zheng Wang (**zhenwang@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. 10th Street
Bloomington, IN 47405 USA

Abstract

A new approach to game theory based on quantum strategies is used to explain some paradoxical phenomena of human choice behavior. Quantum strategies were originally used to explain the fact that humans prefer to cooperate rather than defect in a Prisoner's Dilemma (PD) game. Here we develop a quantum model for the disjunction effect. This refers to a paradox in which (a) a player prefers to defect when the player knows that an opponent will defect, and (b) the player also prefers to defect when the player knows that an opponent will cooperate, but (c) the player reverses preference and cooperates when the opponent's action is unknown. New experimental findings on the disjunction effect are reported, and a quantum explanation for the findings is presented. The quantum model is also compared to traditional information processing models.

Keywords: quantum model; disjunction effect; Prisoner's Dilemma

Quantum Information Processing

Human reasoning and decision making involves a great deal of vagueness, uncertainty, and conflict. How best to model these characteristics is a fundamental question for information processing theories of cognition. In this paper, we examine a quantum computing approach to this problem (see Nielsen & Chuang, 2000).

Consider for example, the decision whether to cooperate or compete with another business on some high tech venture. For example, this other business may have some technical skills that are needed for success. Suppose this decision also depends on whether the other business is trustworthy or untrustworthy. According to a quantum approach, prior to expressing a decision, the decision maker is in a superposition state in which all of the combinations of beliefs about trustworthiness and preferences about cooperation have some potential to be observed. This idea alone is not terribly interesting because any classic information processing theory could also adopt a similar representation. What is interesting is the uniquely quantum idea that possibilities can interfere with each other as if they exist simultaneously in the mental state. In particular, according to quantum theory, the *joint* probability of believing the other business is trustworthy and deciding to cooperate can be *greater* than the marginal probability of deciding to cooperate (which is actually the union of the trustworthy and untrustworthy possibilities). In other words, the probability of the disjunction can fall below the probability of a component event, which is a violation of the OR rule within classic probability theory. Has such a violation ever been empirically observed?

Disjunction Effects

Consider a PD game in which there are two players, you versus other, and each player has two actions: cooperate or compete. An example payoff matrix for each player, conditioned on each pair of actions, is shown in Table 1.

Table 1: Example PD Game.

	You Compete	You Cooperate
Other Competes	You: 10 Other: 10	You: 5 Other: 25
Other Cooperates	You: 25 Other: 5	You: 20 Other: 20

In the standard version of the game, hereafter referred to as the unknown condition, the players simultaneously select an action without knowledge of the opponent's selection. Two new manipulations are used to examine the disjunction effect: In one case, you are initially informed that the other player has chosen to compete; and in another case, you are initially informed that the other player has chosen to cooperate. This manipulation is designed to test the 'sure thing' principle that lies at the foundation of utility theory (Savage, 1954): If you prefer to compete knowing that your opponent will compete *and* you prefer to compete knowing that your opponent will cooperate, then you should prefer to compete even when you do not know your opponents choice.

Shafir and Tversky (1992) found that players frequently violated the sure thing principle – many players chose to compete knowing that the other player competed, and they also chose to compete knowing that the other player chose to cooperate, but they cooperated when they did not know the choice of the other player. See Croson (1999) and Li and Taplan (2002) for replications and extensions.

The disjunction effect also rules out a simple yet important information processing model for this task.

Suppose that during the unknown condition, there are two possible states of thought that you can entertain about your opponent. There is some probability p of thinking 'the other will compete' and there is some complementary probability $(1-p)$ of thinking 'the other will cooperate.' There are also two actions that you can take, compete or cooperate, conditioned on your state of belief. Accordingly, the probability that you choose to compete should equal the weighted average: $p \cdot Pr$[you compete | other competes] + $(1-p) \cdot Pr$[you compete | other cooperates]. This average can never fall outside the range defined by the two extreme probabilities of Pr[you compete | other competes] and Pr[you compete | other cooperates]. However, Shafir and Tversky (1992) reported a competition rate equal to 63% for the unknown state, which fell below the range defined by the two known states: 97% when the other player was known to compete, and 84% when the other player was known to cooperate. This finding contradicts the 'two state' information processing model.

One can save the information processing model by assuming the existence of three mental states: one is a 'known to compete' state, another is a 'known to cooperate' state, and a third is a 'confused' state that is entered when the opponent's action is unknown. Perhaps you tend to randomly guess more whenever you enter the 'confused' state. This could lower the probability of competing during the unknown condition down toward chance (50%), falling below either of the known conditions.

Although the 'three state' model can lower the rate under the 'confused' state toward chance, it cannot lower the rate systematically below chance. Contrary to this prediction, Tversky and Shafir (1992) found a systematic reduction far below chance using a gambling paradigm to study the disjunction effect. In this experiment, you are presented with two possible plays of a gamble that is equally likely to win \$200 or lose \$100. You are instructed that the first play has completed, and now you are faced with the possibility of another play. Tversky and Shafir (1992) found that 69% chose to play again after a known win, 59% chose to play again after a known loss, but only 36% chose to play when the outcome of the first play was unknown. The 'three state' model cannot explain the fact that gambling rate dropped far below the chance rate of 50% during the unknown condition.

New Experiment

To our knowledge, there have been only two replications and extensions of the disjunction effect, and both used the PD game (Croson, 1999; Li & Taplan, 2002). The original study by Tversky and Shafir (1992) and the replication by Li and Taplan (2002) involved deception -- each human actually played against a computer agent. The replication by Croson (1999) used humans playing against humans. The results from the original study and the two replications are compared in Table 2 below.[1] As can be seen, the results from these replications are not as convincing as the original results, as they reveal only minor violations of the predictions of the simple 'two state' model.

Table 2: Summary of Original and Replicated Findings.

Study	Compete	Cooperate	Unknown
Shafir	97	84	63
Croson	67	32	30
Li- Taplan	83	66	60
Matthew	91	84	66
Model	88	81	69

Therefore we decided to conduct another replication. But this time, we conducted a new extension in which humans played against computer agents and the humans were truthfully informed. The instructions were almost identical to those used by Shafir and Tversky (1992), except that participants (Ps) were informed that they would be playing against computer agents on a computer network instead of against other humans. Ps were simply told that they would occasionally see bonus trials that would inform them of the agent's decision prior to making their own.

Our Ps were 88 graduate and undergraduate students at Indiana University who participated for monetary rewards; eligible students also received one credit toward their introductory psychology experimental requirement. We offered a minimum of five dollars, but Ps could earn up to twenty dollars depending on the outcome of their decisions. Our sample consisted of an approximately equal number of males and females (42 and 46, respectively). All Ps were recruited via e-mail from a database of students who had indicated that they would like to be involved in paid experiments.

Each P played 40 games on a computer, 18 of which were PD games similar to that shown in Table 1, and the other games were fillers (not analyzed). All Ps saw PD games on the same trial numbers. Six different PD matrices were used, and each PD matrix was only allowed three presentations. In this manner, all Ps were assured of seeing each of the six PD matrices once in each condition. The exact orders of the PD matrices, however, were determined in a quasi-random fashion. The program always showed the standard form of a matrix the first time that matrix appeared, and then it randomly chose between cooperating and competing the second time (while displaying, of course, the corresponding notification). The third time a matrix was presented, the program chose to compete if it had cooperated the second time or to cooperate if it had competed the second time. In sum, there were six PD matrices and each appeared three times for each of the eighty-eight Ps; we therefore had 528 triads on which to run analyses (a triad consisted of the three cases of each matrix).

[1] The results for Croson were averaged across the symmetric and asymmetric PD game conditions.

Our results are posted in Table 2 in the row labeled Matthew.

The effect of each condition on individual competition rates was first analyzed using a repeated measures ANOVA. Gender was also analyzed, but there was no significant effect. Even after correcting for a slight violation of sphericity, it was clear that the conditions led to significant differences between the rates, $F(1.761, 153.223) = 50.070$, $p<0.001$. This provided the clearance to perform protected paired t-tests. The first such t-test, between the unknown condition and the known-competition condition, produced a significant result, $t(87) = -9.082$, $p<0.001$. Next, we compared the unknown condition to the known-cooperation condition and again found significance, $t(87) = -6.278$, $p<0.001$. Finally, the known-competition and known-cooperation conditions were compared, and significance was again observed, $t(87) = 3.485$, $p<0.001$.

Quantum Strategies

To construct a quantum model for this task, we postulate that there are two states of beliefs about your opponents action, defect versus cooperate. Additionally, there are two states of action for you to take, again defect (which corresponds to competing but allows clearer notation in the coming equations) versus cooperate. We assume that a person can simultaneously consider beliefs and actions which then produces four basis states denoted $\{|DD\rangle, |DC\rangle, |CD\rangle, |CC\rangle\}$. For example, $|DC\rangle$ represents the case where you *simultaneously* believe that the opponent will defect but you *intend* to act cooperatively.

The state of the cognitive system (that is, the part needed for modeling this task) is represented by a quantum state vector. In this case, the state is represented by a 4×1 column vector $\psi = [\psi_{DD}, \psi_{DC}, \psi_{CD}, \psi_{CC}]$. According to quantum theory, the state vector is a probability amplitude distribution across the basis states. For example, ψ_{DC} represents the probability amplitude of the quantum system being observed in state $|DC\rangle$. The probability of observing this state is $|\psi_{DC}|^2$. The state vector must be unit length to guarantee that the state probabilities sum to unity. [2]

The state of the cognitive system is changed by thoughts generated from information in the environment. In terms of the quantum model, a thought is represented by a quantum operator, denoted U, which changes the state from one vector ψ to another $\varphi = U \cdot \psi$. For this application, the quantum operator is represented by a 4×4 unitary matrix U with the property $U^\dagger U = I$, where I is the identity matrix. The matrix U must be unitary in order to preserve the unit length property of the state vector.

The initial state vector represents the state of the cognitive system at the beginning of each trial. This initial state is changed by information given to the P. If the player is informed that the opponent will defect, then an operator is applied to transform the initial state into one that has $\psi_{CD} = \psi_{CC} = 0$ producing $\psi_D = [\alpha_D, \beta_D, 0, 0]$, where $\beta_D^2 = 1 - \alpha_D^2$. If the player is informed that the opponent will cooperate, then another operator is applied that transforms the initial state into one that has $\psi_{DD} = \psi_{DC} = 0$ to produce $\psi_C = [0, 0, \alpha_C, \beta_C]$, where $\beta_C^2 = 1 - \alpha_C^2$. In the unknown case, an operator is applied which produces a superposition state $\psi_U = \sqrt{½} \cdot \psi_D + \sqrt{½} \cdot \psi_C$. The interpretation of this state will be treated in the discussion.

To select a strategy, the player must evaluate the payoffs of the actions. Thus the state ψ is processed by a quantum operator U_t for some period of time t which transforms the previous state into a final state $\varphi = U_t \cdot \psi = [\varphi_{DD}, \varphi_{DC}, \varphi_{CD}, \varphi_{CC}]$. Finally, the observed probability of choosing to defect is given by $|\varphi_{DD}|^2 + |\varphi_{CD}|^2$.

The quantum strategy U_t can be constructed from a Hamiltonian matrix H as follows: $U_t = exp(-i \cdot t \cdot H)$. (This uses a complex matrix exponential function, which is available in MatLab or Mathematica). Here $i = \sqrt{-1}$ and this factor is required to guarantee that U_t is unitary. The processing time parameter, t, is a free parameter in the model, but it can be manipulated by deadline pressure.

In general, the Hamiltonian H must be Hermitian ($H = H^\dagger$) to guarantee that U_t is unitary. For this application, the Hamiltonian is a 4×4 matrix with elements $h_{ij} = h_{ji}^*$.

We use the simplest possible Hamiltonian that can explain the results: $h_{11} = h_{33} = \mu$, and $h_{41} = h_{14} = -1$, where μ is a free parameter. The parameter μ depends on the payoff advantages for defecting as compared to cooperating. This is designed to build up probability amplitude for the defect actions, $|DD\rangle$ and $|CD\rangle$. Setting $h_{41} = h_{14} = -1$ allows transfer of probability amplitude between the correlated states $|DD\rangle$ and $|CC\rangle$. This is designed to coordinate beliefs and actions.

In sum, the model has 4 parameters: α_D and α_C are used to determine the initial states depending on knowledge of the opponents play, and t and μ are used to construct the quantum strategy for taking action. For example, if we set $\alpha_D = 1$, $\alpha_C = .7853$, $t = .5291$, $\mu = 5.3060$, then the model produces the probabilities shown in the bottom row of Table 2, which accurately reproduces the findings of the present study (compare to the row labeled Matthew in Table 2). Alternatively, if we set $\alpha_D = 1$, $\alpha_C = .4061$, $t = .8614$, $\mu = 2.3479$, then the model predicts .69, .59, .36, which perfectly reproduces the Tversky and Shafir (1992) gambling results (here we associate the defect strategy with choosing to play the gamble).

Interference Effects

So, how does the quantum model produce this disjunction effect? Recall that the 'two state' information processing model fails because it must predict that the defection rate for the unknown condition is an average of the rates for the two known conditions. The quantum model violates this property because *interference* effects occur under the unknown condition.

Define M as a matrix with the first row equal to [1 0 0 0] and the second row equal to [0 0 1 0]. The product $\phi = M \cdot \varphi$ produces a 2×1 vector that represents the projection of the quantum state onto the bases that lead one to choose to defect. The squared length, $|\phi|^2 = \phi^\dagger \phi = |\varphi_{DD}|^2 + |\varphi_{CD}|^2$, gives the probability of defection.

If the opponent is known to defect, then we obtain $\phi_D^\dagger \phi_D = (M\varphi_D)^\dagger (M\varphi_D) = (MU_t\psi_D)^\dagger (MU_t\psi_D)$, and if the opponent is

[2] See Neilsen and Chuang (2000) for a review of quantum computing principles.

known to cooperate, then we obtain $\phi_C^\dagger\phi_C = (M\varphi_C)^\dagger(M\varphi_C) = (MU_t\psi_C)^\dagger(MU_t\psi_C)$. During the unknown condition, we obtain $\phi_U^\dagger\phi_U = (M\varphi_U)^\dagger(M\varphi_U) = (MU_t\psi_U)^\dagger(MU_t\psi_U) = [MU_t(\psi_D+\psi_C)]^\dagger[MU_t(\psi_D+\psi_C)]/2 = (\phi_D+\phi_C)^\dagger(\phi_D+\phi_C)/2 = (\phi_D^\dagger\phi_D + \phi_C^\dagger\phi_C)/2 + \phi_D^\dagger\phi_C$.

The latter term $\phi_D^\dagger\phi_C$ is called the interference term. If it is zero, then the quantum model reduces to the traditional 'two state' information processing model, and consequently it fails to account for the disjunction effect. However, if the interference term is negative, then it can reduce the defection rate during the unknown condition below the rates for the known condition. In sum, the superposition state that is generated by the unknown information condition is required to produce the interference effect. But it is not sufficient.

Why is the interference term negative for this quantum model? This is generated by the coordinating link $h_{14} = h_{41} = -1$ in the Hamiltonian, which is used to generate the quantum strategy U_t. This causes the states of the quantum system to become entangled. If this link was turned off, by setting $h_{14} = h_{41} = 0$, then the interference effect disappears. This is true even when the unknown condition produces a superposition state. In conclusion, the combination of superposition and entanglement are required to explain the disjunction effect.

Quantum States and Quantum Strategies

This paper accomplishes two goals. One is to report the results for a new experiment that successfully replicated and extended previous empirical findings concerning the disjunction effect using a PD game. The second is to provide an explanation for the disjunction effect derived from quantum game theory. Our theoretical analyses demonstrate the sufficiency of the quantum model for reproducing the disjunction effect. The 'two state' and 'three state' information processing models fail to do this. We do not have enough data to test the quantum model rigorously at this point. New experiments that include new conditions are being designed for this purpose.

What else can the quantum model predict about the disjunction effect? There are three factors that we can manipulate and use to generate new predictions from the quantum model. One is information about the likelihoods of the opponent's actions, which should affect the superposition state. A second is the payoff matrix of the game, which should affect the Hamiltonian. The third is deadline time pressure, which should affect the processing time parameter. These new manipulations can be used to provide more stringent tests of the model.

What does it mean to be in a superposition state? Consider, for example, belief states. According to a traditional information processing account, the cognitive system may jump from one belief state to another (e.g., jump from one subjective probability to another), but at any given moment, the system is definitely in one particular state of belief. In other words, if we observe the state at some particular moment and find it to be equal to state *X*, then the system must have been in state *X*, immediately before we took this measurement. According to quantum theory, this concept of state is entirely wrong! Immediately before the measurement of state *X*, the system was not exactly located at any particular belief state; instead all belief states coexisted in parallel. For example, a person may have a tendency to believe simultaneously that a defendant is surely guilty and surely not guilty, and this is not the same as believing the defendant is moderately guilty.

How can the quantum model account for the disjunction effect? Shafir and Tversky (1992) offered a possible explanation that may be analogous to entanglement: quasi-magical thinking. Suppose Ps consider that their opponent is engaged in a thought process much like their own (this may apply to the computer agents, also - see Nass & Moon, 2000). No matter what option the Ps consider, they may feel that their like-minded opponents will also consider the same. Because mutual cooperation is more rewarding than mutual competition, Ps are motivated to cooperate so long as they can also reason that their opponents will do the same. Cooperation is only to be preferred however, when mutual cooperation is compared to mutual competition, and Shafir and Tversky point out that the tendency to cooperate should decline when this comparison is no longer relevant (e.g., when mutual competition is impossible because the opponent has already cooperated).

The superposition native to the quantum model is due to the uncertainty over which state is actually relevant during deliberations. Shafir and Tversky state that people may be reluctant to regard probabilistic events as fact and therefore do not explore the resulting paths sufficiently. We offer the view that people have no problem considering probabilistic events as fact but do hesitate to discard their logical opposites under conditions of uncertainty. That is, our Ps may have been perfectly fine believing that their opponents would compete but did not see this as inconsistent with the belief that the opponent would cooperate. We liken this to the familiar state of dreaming in that what goes by unquestioned in our dreams (when everything, so to speak, is superposed) may often present as markedly impossible upon conscious contemplation.

How does processing in quantum systems relate to processing in production rule models or connectionist networks? Similar to production rule models, quantum operators can be programmed to perform sequences of transformations of if-then type using what are called controlled U gates (see Nielsen & Chuang, 2000). Similar to connectionist networks, quantum operators function as content addressable parallel processors that map fuzzy distributed inputs into fuzzy distributed outputs.

What is the neural basis for quantum states and quantum operations? Quantum dynamics are based on wave equations that may provide a good mathematical foundation for modeling brain waves (see Pribram, 1993). Interference effects may be produced by interactions among brain waves.

Finally, are there any other applications of quantum models in cognitive science? Eisert, Wilkens, and Lewenstein (1999, see also Piotrowski & Sladkowski, 2003) opened up a new approach to game theory by introducing quantum principles for selecting mixed or probabilistic strategies. When quantum strategies are permitted, the paradox of the PD game disappears, and cooperation rather than defection emerges as an equilibrium strategy. Bordely

(1998) used quantum probability rules to explain paradoxes in human probability judgments. Mogiliansky, Zamir, and Zwirn (2004) used non commutative measurement operators to explain order effects on human judgments. Gabora and Aerts (2002) developed a quantum theory of concepts. Finally, Busemeyer, Wang, and Townsend (in press) provide a quantum approach to dynamic decisions.

Acknowledgments

This research was supported by NIMH R01 MH068346 to the first author. Thanks to Eric Dimperio for helpful discussions about this paper.

References

Busemeyer, J. R., Wang, Z. & Townsend, J. T. (in press) Quantum dynamics of human decision making. Journal of Mathematical Psychology.

Bordley, R. F. (1998) Quantum mechanical and human violations of compound probability principles: Toward a generalized Heisenberg uncertainty principle. *Operations Research, 46,* 923-926.

Croson, R. (1999). The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes, 80 (2),* 118-133.

Eisert, J., Wilkens, M., & Lewenstein, M. (1999) Quantum games and quantum strategies. *Physical Review Letters, 83,* 3077-3080.

Gabora, L. & Aerts, D. (2002) Contextualizing concepts using a mathematical generalization of the Quantum formalism. *Journal of Experimental and Theoretical Artificial Intelligence, 14*, 327-358.

Li, S., & Taplin, J. (2002) Examining whether there is a disjunction effect in prisoner's dilemma games. *Chinese Journal of Psychology, 44 (1),* 25-46.

Mogiliansky, A. L., Zamir, S., & Zwirn, H. (2004) Type indeterminacy: A model of the KT (Kahneman Tversky) - man. Paper presented at Foundations Utility Risk and Decision Theory XI, Paris, France.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: social responses to computers. *Journal of Social Issues, 56 (1)*, 81-103.

Nielsen, M. A. & Chuang, I. L. (2000) *Quantum computation and Quantum information*. Cambridge, UK: Cambridge University Press.

Piotrowski, E. W.& Sladkowski, J. (2003) An invitation to quantum game theory. *International Journal of Theoretical Physics*, *42*, 1089.

Pribram, K. H. (1993) *Rethinking neural networks: Quantum fields and biological data*. Hillsdale, N. J: Earlbaum

Shafir, E. & Tversky, A. (1992) Thinking through uncertainty: nonconsequential reasoning and choice. *Cognitive Psychology, 24,* 449-474.

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science*, 3, 305-309