Experimental Economics (2024) 27:743–765
https://doi.org/10.1007/s10683-024-09838-w

Explaining interference effects in prisoner dilemma games

Jerome Busemeyer[1,2] · **Masanari Asano**[2,3] · **Meijuan Lu**[1,2]

Received: 27 November 2023 / Revised: 25 June 2024 / Accepted: 3 July 2024 /
Published online: 23 September 2024

Abstract
This article presents a new approach to understanding strategic decision making inspired by the mathematics of quantum theory. Empirical support for this new approach is based on five different puzzling findings from past work on the prisoner dilemma game including the disjunction effect, the interference of predictions on actions in simultaneous and sequential games, question order effect, and the effects of cheap promises. Eight different quantum models are described, which purport to account for these puzzling findings. The competing models are systematically compared with respect to their capability of accounting for the five empirical findings.

Keywords Prisoner dilemma · Disjunction effect · Interference effect · Question order effect · Quantum cognition

JEL code C7

1 Introduction

Time flies and the years will never return. The short journey of life is just a fleeting moment in front of the boundless universe. But there are some extraordinary people who will draw a dazzling light in the darkness like shooting stars in their short life

✉ Jerome Busemeyer
jbusemey@indiana.edu

Masanari Asano
iasano@fuk.kindai.ac.jp

Meijuan Lu
ludaifu.88@gmail.com

[1] Psychological Brain Sciences, Indiana University, 1101 E 10th Street, Bloomington, IN 47405, USA

[2] Department Information Computer Science, Kindai University, 3-4-1 Kowakae, Higashiōsaka 577-8502, Osaka, Japan

[3] Meng River Health, 1950 East Greyhound Pass, Carmel, IN 46033, USA

trajectory, winning the respect and admiration of everyone! Amnon Rapoport is one of the scientists worthy of our respect and admiration.

Amnon Rapoport had a major influence on the careers of many well known scientists in the field of decision making, including the first author. The thorough and insightful review by Rapoport and Wallsten (1972) was the first author's introduction to the field of behavioral decison making. More personally, Amnon provided generous guidance and advice to the first author that helped start his career. So we are very grateful to have the opportunity to honor him in this special issue. Amnon was always a leader for establishing new directions in strategic decision making. We hope we can make a meaningful contribution to follow along some of his paths.

This article presents a new approach to understanding strategic decision making that arises from the mathematics of quantum theory. On the one hand, quantum physicists, anticipating the development of quantum computers, have developed quantum strategies for games, which provide new equilibrium solutions (Meyer, 1999; Eisert et al., 1999; Piotrowski & Sladkowski, 2002; Santos, 2020; Alonso-Sanz, 2019). On the other hand, quantum cognition researchers (reviewed below) have used quantum theory to better describe how people actually make strategical decisions. As we describe below, the main reason for considering quantum theory is that it provides a simple explanation for some puzzling findings obtained from past empirical research on simple economic games.

2 Five puzzling findings

2.1 Disjunction effect

The prisoner dilemma game is one of the most well known experimental economic games in the field (Rapoport, 1988). Briefly, there are two players, call them Ann and Bill, and each player can choose to defect or cooperate. In a one-shot game, they play each other only once. The payoffs for the PD game are arranged so that no matter what choice Ann makes, Bill is better off defecting, and likewise no matter what choice Bill makes, Ann is better off defecting. According to the Nash equilibrium both players should defect, even though both would be better off cooperating. Contrary to this, players frequently cooperate (Rapoport, 1988) producing better outcomes than mutual defection. This deviation from the Nash equilibrium is commonly explained by additional principles like fairness (Fehr & Schmidt, 1999).

Shafir and Tversky (1992) found something more puzzling than cooperative behavior in the PD game – a result that they called the disjunction effect. Their experiments involved three conditions: (1) In one condition, here called the unknown condition, both players moved simultaneously and remained uncertain about the opponent's decision; (2) in a second condition, here called the known defect condition, before her own move, Ann was informed that Bill defected; and (3) and in a third condition, here called the known cooperate condition, before making her move, Ann was informed that Bill cooperated. Their main finding was that the percentage of cooperation in the unknown condition (37%) was significantly greater than the percentage of cooperation in the case when the opponent was known to defect (3%)

and in the case when the opponent was known to cooperate (16%). Shafir and Tversky (1992) claimed that this disjunction effect reflected a violation of the sure-thing principle, because 25% of participants chose to defect when the opponent defected and also chose to defect when the opponent cooperated, but then switched and chose to cooperate when the opponent's decision was unknown. Busemeyer and Matthews (2006) later replicated these findings: the percentage of cooperation in the unknown condition (34%) was greater than that for the case when the opponent was known to defect (9%) and in the case when the opponent was known to cooperate (16%).

The disjunction effect can also be viewed as a violation of a prediction based on the application of the law of total probability. In the unknown condition, there are only two possible moves that the Bill can take. So when Bill's action is unknown, there is some probability, $p(C_B)$, she predicts Bill cooperates and some probability $p(D_B)$ she predicts Bill defects. If she predicts Bill cooperates, then there is a probability, $p(C_A|C_B)$, she cooperates; likewise, if Bill defects, then there is a probability, $p(C_A|D_B)$, she cooperates. Therefore, the probability Ann cooperates in the unknown condition should equal

$$p_T(C_A) = p(C_B) \cdot p(C_A|C_B) + p(D_B) \cdot p(C_A|D_B) \tag{1}$$

The total probability is a weighted average that must lie in between the two conditional probabilities. Contrary to this prediction, the disjunction effect occurs when the proportion of cooperation in the unknown condition lies above both of the known conditions. In sum, information about the opponent before taking an action reduces the total probability of cooperation.

Shafir and Tversky (1992) explained the results by arguing that the advantages of defection are clear when the opponent's action is known, but these advantages become unclear when the opponent's action is unknown. In the unknown condition, Ann clearly knows that Bill can either defect or cooperate, so why can't Ann consider the reasons for each case? This finding can be viewed in terms of Feynman's path integral ideas: if there are two paths a particle can traverse to reach a target and the path is unknown, then they can interfere with each other and cancel each other out. Thus, the intuition from quantum theory is that somehow the two thoughts interfere with each other to produce no clear thought when the Bill's action is unknown.

2.2 Interference of predictions on actions

Croson (1999) conducted 2 studies investigating the effects of predictions on actions using a simultaneous play PD game. These experiments involved two conditions. In one condition (predict-act condition), the player (Ann) was asked to predict the opponent's (Bill's) move before making her own decision; in the other condition (act-alone condition), Ann simply made a decision without expressing any prediction about Bill. The two experiments differed with respect to their payoff matrices. These experiments provided another type of test of the predictions using the law of total probability. In the act-alone condition, there are still only two predictions that Ann can make – predict Bill cooperates or predict Bill defects. If we apply Eq. 1

to this experiment, $p(C_B)$ is the probability that she predicts Bill will cooperate, $p(C_A|C_B)$ is the conditional probability that Ann cooperates given Bill cooperates, $p(D_B)$ is the probability that she predicts Bill will defect, $p(C_A|D_B)$ is the conditional probability that Ann cooperates given Bill defects. So the total probability that Ann cooperates without expressing her prediction, $p(C_A)$, is expected to equal the left hand side of Eq. 1. The results turned out contrary to this expectation: In the first experiment, 77.5% of the participants in the act-alone chose to cooperate, but this dropped to 55% for the predict-act condition; in a second experiment, 62.5% cooperated in the act-alone condition, which dropped to 42.5% in the predict-act condition. The percentage of cooperative predictions was 45% in the first study and 42% in the second study, and the correlation between predictions and actions was $\phi = 0.48$ in the first study and $\phi = 0.30$ in the second study. In sum, making a prediction before taking an action reduced the total probability of cooperation. Simply asking participants to report their prediction reduced the overall percentage of cooperation.

2.3 Question order effects

Tesar (2020) conducted an experiment that included: (a) like Shafir and Tversky (1992), they manipulated the information about the other opponent's (Bill's) move before the player (e.g., Ann) made her decision; (b) like Croson (1999), they manipulated whether or not the player predicted the opponent's (Bill) before the player (e.g., Ann) made her decision; and (c) a new third manipulation that changed the order so that the player (e.g., Ann) first decided an action, and then predicted the opponent's (Bill's) decision. Tesar (2020) replicated both findings by Shafir and Tversky (1992) and Croson (1999), but also found that the percentage of cooperation was 65% when the player (e.g. Ann) made a decision before prediction, but this percentage was significantly reduced to 42% when the player (Ann) predicted the opponent (Bill) first. The percentage of cooperative predictions was higher, 52%, in the act-predict order as compared to the 39% for the predict-act order, and contingency correlation between predictions and actions was lower, $\phi = 0.26$, in the act-predict order as compared to $\phi = 0.36$ for the predict-act order. In sum, changing the order of predictions and actions changed the probability of predicting cooperation, the probability of acting cooperatively, as well as the correlation between predictions and actions.

2.4 Sequential prisoner dilemma game

Blanco et al. (2014) examined the effects of prediction on actions in a sequential prisoner's dilemma game in which one player (say Ann) moves first and the second player (say Bill) responds to the first players move. If Ann defects, then both players receive a payoff of 10 units. If Ann cooperates, then the payoffs depend on Bill's move: if Bill cooperates both players gain a payoff of 14 units; but if Bill defects, he wins 17 units while Ann receives only 7. The subgame perfect equilibrium is defection for Bill, and therefore defection for Ann. The experiment included three conditions: In the prediction condition, each player predicted how many players (out of a total on nine in a session) would cooperate as second movers; in the feedback group, each player was told the number of others that cooperated as second movers; in the baseline group, no prediction was made nor any feedback provided. The percentage of cooperation on the first move for the prediction condition was was 55%; the percentage of cooperation for the feedback condition was 57%; but the percentage of cooperation for the baseline condition was only 28%. The percentage of players predicted to cooperate in the second stage was 51%, and there was a high, $r = 0.87$, correlation between predictions about second stage and first stage actions. In sum, asking for a prediction in the sequential game increased the probability of cooperation.

2.5 Promise effects

Kvam et al. (2014) examined a version of the PD game that allowed for "cheap talk." Each player played several repetitions with computer agents and each repetition involve a sequence of four stages. During stage 1, they were asked to choose whether or not they were willing to promise agent A that they would cooperate in a future PD game with agent A; during stage 2, they played a standard PD game with a different agent B; during stage 3, they played a standard PD game against agent A that they faced in the first stage; and in stage 4, they played several games with other agents without any promises. Half the computer agents had a reputation for cooperation and half had a reputation for defection. The key comparison was the within participant rate of cooperation to agent B in the second stage (immediately following the promise question) as compared to the unrelated fourth stage (which did not include promises). Once again we can apply Eq. 1, in which case we expect the total probability of cooperating on the second stage (pooled across a promise or no promise on stage 1) should equal the probability of cooperating without any promise being considered (stage 4). Kvam et al. (2014) found that there was little difference between cooperation rates when playing an agent with a defecting reputation: in both cases the rate was very low at 16%. However, when playing against agents with a cooperative reputation, the rate of cooperation was significantly higher for the second stage (47%) as compared to the fourth stage (40%).

2.6 Summary of empirical findings

The empirical results reviewed above all indicate violations of expectations based on the application of the law of total probability to choices in the prisoner dilemma game. There are five main findings. First, the disjunction effect refers to the finding that the probability of cooperation when not knowing the action of the opponent falls below both of the conditional probabilities for the known actions of the opponent. Second, the interference of prediction on action refers to the finding that prediction about an opponent changes the later probability of cooperation (pooled over predictions) compared to action without prediction. Third, the direction of the interference changes across types of games: In the simultaneous play version of the PD game, predictions decrease later cooperation, but in the sequential play version, prediction increase later cooperation. Fourth, the question order effect refers to the finding that cooperation is lower when the predictions are made before actions as compared to the.reverse order. Fifth, asking a player to make a promise to one opponent can affect the player's cooperation to later unrelated opponents.

Regarding the direction of interference effects, Blanco et al. (2014) point out that the difference between their results and Croson (1999) was most likely produced by the difference in best response to the opponent's expected move in the two studies. The dominant strategy in the Croson (1999) experiments was to defect, whereas the best response to the expected number of cooperators in the Blanco et al. (2014) study was to cooperate.

Finally, interference effects, defined as the difference between the total probability of action after making a prediction (pooled across predictions) and the the probability of action when no prediction is made, are not limited to the PD game. These effects have also been reported using gambling tasks (Tversky & Shafir, 1992; Broekaert et al., 2020) and categorization-decision tasks (Busemeyer & Lambert-Mogiliansky, 2009).

3 Previous quantum theory accounts

Quantum models of decision making are based on the mathematics of quantum theory, and they have attracted a growing level of interest recently (for a recent review, see Pothos & Busemeyer, 2022). These models have arisen in part as a response to the empirical challenges faced by "rational" decision-making models, which suggest that human behaviour does not align well with classical probability theory.

One the one hand, classical (Kolmogorov, 1950) probability theory is based on assigning probabilities to events defined as subsets of a universal set (sample space). These subsets obey a Boolean algebra (or more generally a σ field) that satisfy the properties of commutativity and distributivity. These two properties make it possible to derive the classical law of total probability. Quantum (Von Neumann, 1955) probability theory is based on assigning probabilities to events defined as subspaces of a vector (Hilbert) space. These subspaces only form a partial Boolean algebra that do not necessarily satisfy the properties of commutativity and distributivity. For this reason, quantum probabilities do not necessarily obey the law of total probability (Hughes, 1989). These violations of the law of total probability allow quantum theory to provide a natural explanation for the empirical interference effects reviewed above. In this sense, quantum theory provides a generalized probability theory that relaxes some of the axioms of classical probability theory. This generalization of probability theory allows it to account for puzzling findings including context effects (e.g., Bruza et al., 2023), interference effects (e.g., Kvam, Pleskac, Yu, & Busemeyer, 2015) and constructive judgements (e.g., White, Pothos, & Busemeyer, 2014).

Then why quantum theory and not some other probability theory? The answer is a famous theorem by Gleason (1957): quantum probabilities are the only way to assign probabilities to subspaces that form an additive measure for vector spaces with dimensions greater than 2. See (Busemeyer & Bruza, 2012; Khrennikov, 2010) for introductions to quantum probability theory applied to human decision making.

There are now at least 9 different quantum cognition models that have been developed to account for the puzzling disjunction and interference effects described above (Pothos & Busemeyer, 2009; Kvam et al., 2014; Yukalov & Sornette, 2014; Asano et al., 2011a; Denolf et al., 2016; Martínez-Martínez & Sánchez-Burillo, 2016; Tesar, 2020; Tanaka et al., 2022). Below we summarize and critically evaluate these models, beginning with the simplest versions and progressing to more complex versions. The basic principles of quantum theory are introduced within each application.[1]

3.1 Tesar model

Tesar (2020) proposed a simple two dimensional quantum model for the disjunction effect in the PD game based on the following ideas. First consider the condition in which the player (e.g., Ann) predicts her opponent's (Bill's) action, and then makes her own decision.

Ann's predictions are assumed to be based on a belief state that has a potential α_D for predicting defection and another potential α_C for predicting cooperation. In general, these potentials can be complex numbers, but their squared magnitudes must sum to one. These two potentials form a vector $\psi_U = \begin{bmatrix} \alpha_D \\ \alpha_C \end{bmatrix}$. The vector ψ_U represents Ann's initial state when Bill's actions are unknown. To compute the probability that Ann predicts Bill defects, denoted $p(D_B)$, we first define a projector $P_D = diag\begin{bmatrix} 1 & 0 \end{bmatrix}$ that is used to pick the first coordinate in ψ_U, and then compute the probability from the squared projection: $p(D_B) = \left\| P_D \cdot \psi_U \right\|^2 = \left| \alpha_D \right|^2$. The probability that Ann predicts Bill cooperates is then computed from $p(C_B) = \left\| P_C \cdot \psi_U \right\|^2 = \left| \alpha_C \right|^2$, where $P_C = I - P_D$ and I is the identity matrix.

If Ann eventually does predict (or is told) that Bill defects, then her belief state is updated according to $\psi_D = \frac{P_D \cdot \psi_U}{\sqrt{p(D_B)}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Likewise if Ann eventually does predict (or is told) that Bill cooperates, then her belief state is updated according to $\psi_C = \frac{P_C \cdot \psi_U}{\sqrt{p(D_C)}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

To evaluate her own action decision, Ann rotates her state from a prediction point of view to an action point of view based using a 2×2 unitary "rotation" matrix $U = \begin{bmatrix} u_{DD} & u_{DC} \\ u_{CD} & u_{CC} \end{bmatrix}$, where for example u_{DC} represents the transition to action state D from prediction state C. This unitary matrix is a complex valued, orthonormal matrix, which is length preserving. Ann's new state after evaluation then equals $U \cdot \psi_j$, $j = C, D, U$.

[1] Knowledge of quantum physics is not needed. The models discussed below are all based on finite dimensional spaces. Only a background in matrix algebra is required.

The projectors for Ann deciding to be cooperative versus defective is represented by the same diagonal matrices as before, $P_C = diag\begin{bmatrix} 0 & 1 \end{bmatrix}$, and $P_D = diag\begin{bmatrix} 1 & 0 \end{bmatrix}$, respectively. The probability that Ann chooses the action to cooperate, given that she predicts Bill cooperates (also in the case that she is told Bill cooperated) equals $p(C_A|C_B) = \left\|(U^\dagger \cdot P_C \cdot U) \cdot \psi_C\right\|^2 = \left|u_{CC}\right|^2$, and her probability to cooperate given she predicts Bill defects equals $p(C_A|D_B) = \left\|(U^\dagger \cdot P_C \cdot U) \cdot \psi_D\right\|^2 = \left|u_{CD}\right|^2$.

Based on the above computations, the probability that Ann predicts Bill defects and she decides to cooperate equals

$$\begin{aligned} p(D_B, C_A) &= p(D_B) \cdot p(C_A|D_B) \\ &= \left\|(U^\dagger \cdot P_C \cdot U) \cdot P_D \cdot \psi_U\right\|^2 \end{aligned} \tag{2}$$

The effect of question order is produced by reversing the processing order

$$p(C_A, D_B) = \left\|P_D \cdot (U^\dagger \cdot P_C \cdot U) \cdot \psi_U\right\|^2 \tag{3}$$

Finally, consider the case when Ann directly decides her own action when she does not predict (or when she remains uncertain about Bill's action). In this case, she starts from the initial superposition state over predictions, ψ_U. Then her probability to cooperate for the unknown (or act - only) condition equals

$$\begin{aligned} p(C_A|U) &= \left\|P_C \cdot U \cdot \psi_U\right\|^2 \\ &= \left|u_{CD} \cdot \alpha_D + u_{CC} \cdot \alpha_C\right|^2 \\ &= \left|u_{CD} \cdot \alpha_D\right|^2 + \left|u_{CC} \cdot \alpha_C\right|^2 + 2 \cdot Real\left[u_{CC}^* \cdot \alpha_C^* \cdot u_{CD} \cdot \alpha_D\right] \\ &= p(D_B) \cdot p(C_A|D_B) + p(C_B) \cdot p(C_A|C_B) + Int_C. \end{aligned} \tag{4}$$

Note that $P_C^\dagger P_C + P_D^\dagger P_D = I$ (the identity), which guarantees that $p(C_A|U) + p(D_A|U) = 1$. The term, Int_C, is the interference. If $Int_C = 0$, then $p(C_A)$ obeys the law of total probability (Eq. 1). Positive Int_C increases cooperation for $p(C_A)$ above that predicted by the total probability; negative Int_C decreases cooperation for $p(C_A)$ compared to that predicted by the total probability. Choosing a unitary matrix U to produce positive interference can then be used to account for the observed disjunction effect.

Although this model contains several unknown parameters, it does make two very strong a priori predictions. First of all, the unitary nature of the rotation matrix U requires double stochasticity: the squared lengths of the rows equal one and so do the squared lengths of the columns. In the case of a 2×2 doubly stochastic matrix, this implies $p(C_A|D_B) = p(D_A|C_B)$. Second the reversed operations used to produce the question order effect require satisfaction of the QQ equality (see Busemeyer & Bruza, 2012, Ch. 3): $QQ = p(C_A, D_B) + p(D_A, C_B) = p(D_B, C_A) + p(C_B, D_A)$. Both of these predictions were supported in the data reported by Tesar (2020). However, the double stochasticity condition is strongly violated by the results of Shafir and Tversky (1992) where they found $p(C_A|D_B) = 0.03$ and $p(D_A|C_B) = 0.83$.

3.2 Yukalov and Sornette model

Yukalov and Sornette (2011, 2014) proposed a four dimensional model to explain the disjunction effect summarized as follows. They begin by assuming a (e.g., Ann's) cognitive state represented by a 4×1 unit length vector[2]

$$\psi = \left[\psi^*_{DD}\ \psi^*_{DC}\ \psi^*_{CD}\ \psi^*_{CC}\right]^\dagger. \tag{5}$$

The first index stands for the Ann's prediction, and the second index stands for Ann's action. For example, ψ_{CD} is the potential to predict Bill cooperates but Ann defects.

The key idea of this model is that Ann's decision to cooperate in the unknown (or the act-alone) condition is not based only on the cognitive state ψ. Instead it is generated from another vector $\pi_C = \frac{1}{\sqrt{K}}\left[\,0\ \phi^*_1\ 0\ \phi^*_2\,\right]^\dagger$, called the prospect for cooperation. The probability to cooperate for the unknown (or act-only) condition is computed from the squared magnitude of the inner product

$$\begin{aligned} p(C_A|U) &= \left|\pi^\dagger_C \cdot \psi\right|^2 \\ &= \left|\phi^*_1 \cdot \psi_{DC} + \phi^*_2 \cdot \psi_{CC}\right|^2 \\ &= \left|\phi^*_1 \cdot \psi_{DC}\right|^2 + \left|\phi^*_2 \cdot \psi_{CC}\right|^2 \\ &\quad + 2 \cdot Real\left(\phi^*_1 \cdot \psi_{DC} \cdot \phi_2 \cdot \psi^*_{CC}\right) \\ &= p_{DC} + p_{CC} + Int_C \end{aligned} \tag{6}$$

with $p_{DC} = \left|\phi^*_1 \cdot \psi_{DC}\right|^2$ and $p_{CC} = \left|\phi^*_2 \cdot \psi_{CC}\right|^2$.

Ann's decision to defect in the unknown (or act-only) condition is generated from the cognitive state by using another vector called the prospect for defection defined by $\pi_D = \frac{1}{\sqrt{K}}\left[\,\varphi^*_1\ 0\ \varphi^*_2\ 0\,\right]^\dagger$. The probability to defect during the unknown (or act-alone) condition is then computed from the squared magnitude of the inner product

[2] In general the coefficients may be complex valued. The asterisks represents conjugation and the dagger represents conjugate transpose. The vector is written this conjugated way to avoid writing a four row column vector.

$$\begin{aligned} p(D_A|U) &= \left|\pi_D^{\dagger}\cdot\psi\right|^2 \\ &= \left|\varphi_1^*\cdot\psi_{DD}+\varphi_2^*\cdot\psi_{CD}\right|^2 \\ &= \left|\varphi_1^*\cdot\psi_{DD}\right|^2+\left|\varphi_2^*\cdot\psi_{CD}\right|^2 \\ &+2\cdot Real\left(\varphi_1^*\cdot\psi_{DD}\cdot\varphi_2\cdot\psi_{CD}^*\right) \\ &= p_{DD}+p_{CD}+Int_D \end{aligned} \tag{7}$$

where we have defined $p_{DD}=\left|\varphi_1^*\cdot\psi_{DD}\right|^2$, $p_{CD}=\left|\varphi_2^*\cdot\psi_{CD}\right|^2$.

The normalization factor K is needed to satisfy $p(C_A|U)+p(D_A|U)=1$; i.e., K is selected to satisfy $\left|\pi_C^{\dagger}\cdot\psi\right|^2+\left|\pi_D^{\dagger}\cdot\psi\right|^2=1$. Note that the normalization factor K depends on the state ψ. Again the empirical fact that $p(C_A|U)+p(D_A|U)=1$ implies that $Int_C+Int_D=0$, which then implies that $p_{CC}+p_{DC}+p_{CD}+p_{DD}=1$.

Thus, the distribution $(p_{CC}, p_{DC}, p_{CD}, p_{DD})$ can be interpreted as a classical joint probability distribution. These joint probabilities are factored into marginal and conditional probabilities so that $p_{ij}=p(i)\cdot p(j|i)$, where $p(i)$ is interpreted as the probability that Bill takes action i, and $p(j|i)$ is interpreted as the probability that Ann takes action j given Bill takes action i. This model is then applied to the results by setting $p(i)$ equal to the observed proportion that the opponent (e.g., Bill) is predicted (or is told) to take action i, and setting $p(j|i)$ equal to the observed proportions of the player's (e.g. Ann's) action given the opponent's action.

Finally, in order to make a prediction about the probability to cooperate when Bill's action is unknown, Yukalov and Sornette (2011) use what they call the interference quarter law: assuming a flat prior across interference terms, the expected interference is $\pm\frac{1}{4}$, and only the sign of the interference needs to be determined. So finally, the probability to cooperate in the unknown condition equals

$$p(C_A|U)=p(C)\cdot p(C|C)+p(D)\cdot p(C|D)\pm\frac{1}{4},$$

If $Int_C=+.25$,, then the model produces cooperation for $p(C_A|U)$ above the total probability, and if $Int_C=-.25$,, then the model produces cooperation for $p(C_A|U)$ below the total probability. For example, to account for the Shafir and Tversky (1992) results, the interference needs to be positive.

The use of the quarter law makes this theory highly testable with respect to predicting the size of the interference. However, there are several issues with this model. One problem is that the model does not attempt to predict the marginal or conditional response probabilities, and instead it simply estimates them from the data. Another problem is that the model does not explain the change in the sign of the interference across studies. A third is that the interference term is theoretically bounded by the observed choice probabilities (see Appendix A), but the quarter law ignores this bound. In fact, a violation of the bound occurs when applying the quarter law to the disjunction effects Shafir and Tversky (1992); Busemeyer and Matthews (2006) (see Appendix A). A fourth problem is that the interference is not consistently well approximated by the quarter law. For example, considering the study by Kvam et al. (2014), the empirically observed interference equals .07 which is far below the quarter law. Finally, the model has not been developed to account for question order effects.

3.3 Pothos and Busemeyer model

Pothos and Busemeyer (2009) also proposed a 4 dimensional quantum model. The player's (e.g., Ann's) cognitive state is again represented as a 4 dimensional unit length vector generally described by Eq. 5. However, Pothos and Busemeyer (2009) assumed the player (e.g. Ann) starts out the game in an initial belief state, $\psi_U = \frac{1}{\sqrt{2}}\left[\begin{array}{cccc} \phi_1^* & \phi_1^* & \phi_2^* & \phi_2^* \end{array}\right]^\dagger$. The parameter ϕ_1 represents the potential for Ann to predict Bill defects, and ϕ_2 is the potential to predict Bill cooperates, and $|\phi_1|^2 + |\phi_2|^2 = 1$. This initial state begins by assigning equal potentials, $\frac{1}{\sqrt{2}}$ for Ann to take the cooperation or defection actions.

The projector for predicting Bill cooperates is represented by a diagonal matrix $P_{PC} = diag\left[\begin{array}{cccc} 0 & 0 & 1 & 1 \end{array}\right]$, which picks out the potentials for predicting Bill cooperates; the projector for predicting Bill defects is the complement $P_{PD} = I - P_{PC}$, where I is the identity. The probability that Ann predicts Bill cooperates then equals $p(C_B) = \left\|P_{PC} \cdot \psi_U\right\|^2 = |\phi_2|^2$ and the probability she predicts Bill defects is $p(D_B) = \left\|P_{PD} \cdot \psi_U\right\|^2 = |\phi_1|^2$.

If Bill is known or predicted to cooperate, then Ann's belief state is projected to $\psi_C = \frac{P_{PC} \cdot \psi}{\sqrt{p(C_D)}} = \frac{1}{\sqrt{2}}\left[\begin{array}{cccc} 0 & 0 & 1 & 1 \end{array}\right]^\dagger$. Likewise, if Bill is known or predicted to defect, then Ann's initial belief state is projected to $\psi_D = \frac{1}{\sqrt{2}}\left[\begin{array}{cccc} 1 & 1 & 0 & 0 \end{array}\right]^\dagger$. If Bill's action is unknown or not predicted, then Ann remains in the initial unknown belief state, ψ_U but also note that $\psi_U = \left(\phi_1 \cdot \psi_C + \phi_2 \cdot \psi_D\right)$.

Before an action can be chosen, Ann needs to evaluate the actions based on the payoffs of the game. This evaluation is represented by a unitary transformation, U, of the current state $\psi_j, j = C, D, U$ to produce a new action state $U \cdot \psi_j$. We describe how U is built later in this subsection.

A decision is generated probabilistically from the new action state by using a projector that picks out the potentials associated with an action. The projector for Ann to cooperate is given by $P_{AC} = diag\left[\begin{array}{cccc} 0 & 1 & 0 & 1 \end{array}\right]$, and the projector for Ann to defect equals $P_{AD} = diag\left[\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}\right]$. The probability for Ann to cooperate then equals the squared magnitude of the projection of the action state on cooperative actions $p(C_A|j) = \left\|P_{AC} \cdot U \cdot \psi_j\right\|^2$, and the probability for Ann to defect equals the squared magnitude of the projection of the action state on defection actions $p(D_A|j) = \left\|P_{AD} \cdot U \cdot \psi_j\right\|^2$. More specifically, if Ann predicted (or was told) Bill cooperated, then her probability to cooperate is $p(C_A|C_B) = \left\|P_{AC} \cdot U \cdot \psi_C\right\|^2$; and if Ann predicted (or was told) Bill defected then her probability to cooperate is $p(C_A|D_B) = \left\|P_{AC} \cdot U \cdot \psi_D\right\|^2$;

If Bill's action remains uncertain, then

$$\begin{aligned}p(C_A|U) &= \left\|P_{AC} \cdot U \cdot \psi_U\right\|^2 \\ &= \left\|P_{AC} \cdot U \cdot \left(\phi_1 \cdot \psi_C + \phi_2 \cdot \psi_D\right)\right\|^2 \\ &= \left\|\phi_1 \cdot P_{AC} \cdot U \cdot \psi_C + \phi_2 \cdot P_{AC} \cdot U \cdot \psi_D\right\|^2 \\ &= \phi_1^2 \cdot \left\|P_{AC} \cdot U \cdot \psi_C\right\|^2 + \phi_2^2 \cdot \left\|P_{AC} \cdot U \cdot \psi_D\right\|^2 \\ &+ Real\left(\psi_C^\dagger \cdot U^\dagger \cdot P_{AC}^\dagger \cdot \phi_1^* \cdot \phi_2 \cdot P_{AC} \cdot U \cdot \psi_D\right) \\ &= \frac{1}{2}p(C_A|C_P) + \frac{1}{2}p(C_A|D_P) + Int_C.\end{aligned} \tag{8}$$

Note that $P_{AC}^\dagger P_{AC} + P_{AD}^\dagger P_{AD} = I$ (the identity), which guarantees that $p(C_A|U) + p(D_A|U) = 1$. Once again the interference can be positive or negative with this model, depending on the unitary transformation U.

Similar to the Tesar model, this model predicts order effects by reversing the order of the projections

$$\begin{aligned}p(D_B, C_A) &= \left\|\left(U^\dagger \cdot P_{AC} \cdot U\right) \cdot P_{PD} \cdot \psi_U\right\|^2 \\ p(C_A, D_B) &= \left\|P_{PD} \cdot \left(U^\dagger \cdot P_{AC} \cdot U\right) \cdot \psi_U\right\|^2.\end{aligned} \tag{9}$$

Pothos and Busemeyer (2009) construct the unitary rotation, U, from the payoffs for each action as follows. The unitary transformation was based on the matrix exponential, $U = exp\left(-i \cdot \frac{\pi}{2} \cdot H\right)$ of a Hamiltonian built from two parts

$$H = H_1 + H_2$$

$$H_1 = \begin{bmatrix} \frac{\mu_1}{\sqrt{1+\mu_1^2}} & 1 & 0 & 0 \\ 1 & \frac{-\mu_1}{\sqrt{1+\mu_1^2}} & 0 & 0 \\ 0 & 0 & \frac{\mu_2}{\sqrt{1+\mu_2^2}} & 1 \\ 0 & 0 & 1 & \frac{-\mu_2}{\sqrt{1+\mu_2^2}} \end{bmatrix}, H_2 = \frac{-\gamma}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

The first part H_1 is determined by the game payoffs. The upper left 2×2 sub-matrix of H_1 rotates the potentials for the actions in the directions determined by the utility parameter μ_1, with μ_1 an increasing function of the advantage for defecting if the opponents defects; the lower right 2×2 sub-matrix of H_1 rotates the potential for the actions in directions determined by the utility parameter μ_2, with μ_2 an increasing function of the advantage for defecting if the opponents cooperates.

The second part H_2 implements what Shafir and Tversky (1992) called "wishful thinking' – Ann wishes that the own strategy is same with Bill's strategy. It coordinates actions with beliefs by increasing the potentials for ψ_{CC}, ψ_{DD} and decreasing the potentials for ψ_{CD}, ψ_{DC}. The second part of the Hamiltonian is critical for producing interference effects: the sign of the interference effect depends on the sign of γ, and if $\gamma = 0$ then no interference effects can occur Pothos and Busemeyer (2009). This model also accounts for order effects using the same reversal of order of projections as used in the Tesar model. In sum, this model accounts for the interference effects, and it also accounts for question order effects.

Pothos and Busemeyer (2009) applied the model to the disjunction effect using only two parameters. They fixed the initial belief state parameters to $\phi_1 = \phi_2 = \sqrt{.5}$, and they constrained $\mu_1 = \mu_2 = \mu$, because the advantage for defecting was the same amount regardless of whether the opponent cooperated or defected in these studies. Using the estimates $\mu = 0.5263$ and $\gamma = 2.2469$ fit to the experiment by Busemeyer and Matthews (2006) produces the following model predictions: $p(C_A|C_B) = 0.28, p(C_A|D_B) = 0.18, p(C_A|U) = 0.35$. The model is able to produce a disjunction effect, but not as extreme as the results of the Busemeyer and Matthews (2006) experiment.

3.4 Denolf and Martínez model

Denolf et al. (2016) developed a quantum model for the Blanco et al. (2014) experiment. They proposed a simple 2-dimensional model for this experiment. Accordingly, Ann's initial state is a 2×1 column vector $\psi = \left[\, \psi_C \ \psi_D \,\right]^\dagger$. For simplicity, they assumed that these potentials are real valued. The initial state has unit length so that $\psi_D^2 = 1 - \psi_C^2$. If no prediction is made (i.e., the baseline condition), then the probability that Ann cooperates on the first stage simply equals $p(C_A) = \left|\psi_C\right|^2$.

Recall that in the prediction condition of this experiment, the player (e.g., Ann) had to predict how many of the other 9 other players would cooperate. To represent the 10 possible predictions about the number of players, they propose that each prediction corresponds to a projector $P_j = B_j \cdot B_j^\dagger$ for prediction $j = 0, 1, \ldots 9$, where B_j is a unit length 2×1 vector.The two extreme belief states (0, 9) are assumed to be orthogonal $\left(B_0^\dagger \cdot B_9 = 0\right)$. The intermediate belief states are spaced at equal increments in angles between these two extremes. The probability of predicting $j = 0, \ldots, 9$ is determined from the squared length of the projection ψ on vector B_j, which equals $\left\|P_j \cdot \psi\right\|^2 = \left|B_j^\dagger \cdot \psi\right|^2$. However, $\sum_j P_j^\dagger \cdot P_j \neq I$ so that $K = \sum_k \left|B_j^\dagger \cdot \psi\right|^2 \neq 1$ and the squared projections need to be divided by a normalization factor K, which depends on the state ψ, to obtain the probability: $p(j|\psi) = \frac{1}{K} \cdot \left|B_j^\dagger \cdot \psi\right|^2$.

Ann's state after predicting j is reduced to the unit vector B_j. The event representing Ann's decision to cooperate on the first stage is represented by a projector $P_C = A \cdot A^\dagger$, where A is a unit length 2×1 vector. The event representing Ann's decision to defect on the first stage is the complement $P_D = I - P_C$. The probability that Ann decides to cooperate on the first stage after making the prediction j is then equal to the squared length of the projection of her current state B_j on the event representing cooperation, A, which equals $p(C_A|j) = \left\|P_C \cdot B_j\right\|^2 = \left|A^\dagger \cdot B_j\right|^2$. Likewise, the probability that Ann defects after making the prediction j equals $p(D_A|j) = \left\|P_D \cdot B_j\right\|^2$, and note $P_C^\dagger P_C + P_D^\dagger P_D = I$ so that $p(C_A|j) + p(D_A|j) = 1$.

An interference effect occurs because

$$\begin{aligned} \left|\psi_C\right|^2 &\neq \sum_j p(C_A|j) \cdot p(j|\psi) \\ &= \frac{1}{K} \sum_j \left|A^\dagger \cdot B_j\right|^2 \cdot \left|B_j^\dagger \cdot \psi\right|^2. \end{aligned} \tag{10}$$

The difference between the left and right hands sides of Eq. 10 can determine the interference effect.

The Denolf et al. (2016) model includes as free parameters two angles to determine cosines between the belief vector B_0 and ψ and B_0 and A; plus the potential ψ_C in ψ representing the initial state. After fitting these parameters to the data, the model closely reproduce all the results of Blanco et al. (2014). However, this model has never been used to account for the disjunction effect or question order effects.

3.5 Kvam and Mogiliansky model

Kvam et al. (2014) developed a "type indeterminacy" model (Lambert-Mogiliansky et al., 2009) for the effects of promises on playing the PD game. They postulated three types of tendencies that a person could have in a PD game: a tendency T_1 to cooperate, a tendency T_3 to defect, and a moderate tendency T_2 to cooperate if the agent cooperates and defect if the agent defects. A person (e.g., Ann) is not exactly any one of these types, and instead she is superposed over the types. During the first stage, these types have different probabilities to choose to promise or not promise cooperation.

They represent the Ann's state at stage 1 as a 6×1 unit length column vector

$$\psi = \left[\, \psi_1 \cdot \alpha_{P|1} \;\; \psi_1 \cdot \alpha_{N|1} \;\; \psi_2 \cdot \alpha_{P|2} \;\; \psi_2 \cdot \alpha_{N|2} \;\; \psi_3 \cdot \alpha_{P|3} \;\; \psi_3 \cdot \alpha_{N|3} \,\right]^\dagger.$$

(This model is restricted to real values). For example, ψ_2^2 is the probability that Ann becomes a type 2 agent, $\alpha_{P|2}^2$ is the probability that Ann promises if she is a type 2 agent, and $\alpha_{N|2}^2$ is the probability that Ann does not promise if she is a type 2 agent. The potentials of this state vector depend on the reputation (cooperative or not cooperative) of the computer agent.

Let's consider the situation of an agent with a cooperative reputation. In this case, they assume that $\alpha_{P|1} = 1$, $\alpha_{P|2} = 1$, $\alpha_{N|3} = 1$ (for an agent that has a non-cooperative reputation, $\alpha_{P|2} = 0$). So when Ann plays a cooperative agent, denoted here as CA, her initial state is $\psi_{CA} = \left[\, \psi_1 \;\; 0 \;\; \psi_2 \;\; 0 \;\; 0 \;\; \psi_3 \,\right]^\dagger$.

Choosing to promise is represented by the projector $P_P = diag\left[\begin{array}{cccccc}1 & 0 & 1 & 0 & 1 & 0\end{array}\right]$, and choosing not to promise is represented by $P_N = I - P_P$, where I is the identity. The probability that Ann promises to cooperate to a cooperative agent equals $p(P) = \left\|P_P \cdot \psi_{CA}\right\|^2 = \psi_1^2 + \psi_2^2$; and if Ann chooses to promise, then her new conditional state equals $\psi_P = \frac{P_P \cdot \psi_{CA}}{\sqrt{p(P)}}$. The probability that Ann does not promise to a cooperative agent equals $p(N) = \left\|P_N \cdot \psi_{CA}\right\|^2 = \psi_3^2$; and if Ann chooses not to promise, then her new conditional state equals $\psi_N = \frac{P_N \cdot \psi_{CA}}{\sqrt{p(N)}}$. If no promise is made, then Ann remains in the initial state ψ_{CA} for a cooperative agent.

The decision to cooperate or not with agent B in the second stage requires an evaluation about the actions. This evaluation is represented by a 6×6 rotation matrix U. The rotation matrix has only three rotation parameters, because it only operates on the subspace containing potentials (ψ_1, ψ_2, ψ_3). The new state after evaluation is $U \cdot \psi_i$, where $i = P, N, CA$. Using the evaluation basis, the decision to cooperate is represented by a projector $P_C = diag\left[\begin{array}{cccccc}1 & 0 & 1 & 0 & 1 & 0\end{array}\right]$, and the projector for the decision to defect equals $P_D = I - P_C$. The probability that Ann cooperates after promising to cooperate equals $p(C|P) = \left\|P_C \cdot U \cdot \psi_P\right\|^2$; the probability that Ann cooperates after not promising to cooperate equals $p(C|N) = \left\|P_C \cdot U \cdot \psi_N\right\|^2$;

The total probability that Ann's cooperates at stage 2, after choosing to promise or not, then equals $p_T(C) = p(P) \cdot p(C|P) + p(N) \cdot p(C|N)$. The probability that Ann cooperates at stage 4 when there are no promises to make is equal to $p(C) = \left\|P_C \cdot U \cdot \psi_{CA}\right\|^2$, which can be decomposed into the total probability plus interference in the same way as Eq. 8. The interference term then accounts for the effect of promises made of stage 1 to decisions made on stage 4. Although this model was not applied to question order effects, it could be applied in the same way as the Tesar and Pothos-Busemeyer models by reversing the order of projection operations (as in Eq. 9).

Kvam et al. (2014) estimated ψ for each type of agent from the data for each computer agent reputation. The unitary matrix U was derived from three rotation parameters also estimated from the data. These parameters were then used to predict the cooperation data from stages 2 and 4 for each computer agent reputation, which provided an accurate fit the data. However, this model has never been used to account for the Shafir and Tversky (1992) disjunction effect.

3.6 Asano Ohya Tanaka Basieva Khrennikov model

Asano et al. (2011b) and later Asano et al. (2012) proposed a dynamic quantum operations model for the PD game. The state (e.g., Ann's state) regarding each action is initially represented by a 2×1 unit length vector $\psi = \left[\begin{array}{cc}\psi_C & \psi_D\end{array}\right]^\dagger$. This vector is used to form what is called a density matrix in quantum theory, denoted ρ, which is defined by the outer product: $\rho(0) = \psi \cdot \psi^\dagger$. The diagonal values of

ρ, denoted ρ_C for the upper diagonal and ρ_D for the lower diagonal, contain the probabilities for Ann to cooperate and defect, respectively. It is hypothesized that Ann vacillates between thinking about cooperating and defecting, and during this deliberation, the density matrix evolves over time according to the following quantum updating operation (τ represents a small time increment for each update)

$$\begin{aligned}
\rho(t+\tau) &= \Gamma_1(\tau)\cdot\rho(t)\cdot\Gamma_1^\dagger(\tau)+\Gamma_2(\tau)\cdot\rho(t)\cdot\Gamma_2^\dagger(\tau),\\
\Gamma_1(\tau) &= \begin{bmatrix}\sqrt{1-\Delta_D(\tau)} & 0\\ 0 & \sqrt{1-\Delta_C(\tau)}\end{bmatrix},\\
\Gamma_2(\tau) &= \begin{bmatrix}0 & \sqrt{\Delta_C(\tau)}\cdot e^{i\theta_2}\\ \sqrt{\Delta_D(\tau)}\cdot e^{i\theta_1} & 0\end{bmatrix}.
\end{aligned} \tag{11}$$

Note that $\Gamma_1^\dagger\Gamma_1+\Gamma_2^\dagger\Gamma_2=I$, which guarantees that $\rho_C(t)+\rho_D(t)=1$ across time. Assuming $0<\Delta_i<1$, $i=1,2$, then the off diagonal elements of $\rho(t)$ eventually decay toward zero, and the diagonal elements converge toward the asymptote

$$\begin{aligned}
\rho_C &= \frac{\Delta_C}{\Delta_D+\Delta_C}\\
\rho_D &= \frac{\Delta_D}{\Delta_D+\Delta_C}.
\end{aligned}$$

The updating weights, $\Delta_i, i=C,D$ are determined by two amplitudes, μ_C and μ_D as follows:

$$\Delta_C(\tau)=\tau\cdot\frac{|\mu_C|^2}{|\mu_C|^2+|\mu_D|^2},\ \Delta_D(\tau)=\tau\cdot\frac{|\mu_D|^2}{|\mu_C|^2+|\mu_D|^2}.$$

The two amplitudes, μ_C and μ_D, are in turn derived from the beliefs about Bill's moves and the utilities of the payoffs to Ann.

The basic intuition is that Ann imagines the advantages produced by different possible changes. The notation $\mu_{CC\to CD}$ symbolizes the potential for the strategies to change from the *CC* (Bill cooperates, Ann cooperates) to the *CD* (Bill cooperates, Ann defects) pair, and the magnitude of this potential is an increasing function of the payoff advantage to Ann for making this change. Likewise, $\mu_{DC\to DD}$ is the potential for changing from the *CD* to *DD*, $\mu_{DC\to CD}$ is the potential for changing from *DC* to *CD*, and $\mu_{DD\to CC}$ is the potential for changing from the *DD* to the *CC*. Then the potentials μ_D and μ_C are defined as

$$\begin{aligned}
\mu_D &= \alpha_1\cdot\mu_{CC\to CD}+\alpha_2\cdot\mu_{DC\to DD}+\alpha_3\cdot\mu_{DC\to CD}\\
\mu_C &= \alpha_4\cdot u_{DD\to CC.}
\end{aligned} \tag{12}$$

The coefficients, $\alpha_i, i=1,\dots,4$, represents beliefs that these transitions occur.

The disjunction effect can be obtained from this model because the potentials μ_D and μ_C change for known and unknown conditions. In particular, the "wishful thinking" transition $DD\to CC$ is only possible if Ann is uncertain about Bill's action.

The main advantage of this model is that it provides a description of the dynamic evolution of preference during Ann's decision. However, this model has not been developed to account for predictions that a player might make concerning an opponent's play, and consequently, it is unable to account for question order effects.

3.7 Martínez Sánchez-Burillo model

Martínez-Martínez and Sánchez-Burillo (2016) also proposed a dynamic quantum model for the PD game. Once again, they begin by assuming a cognitive state (e.g., Ann's) represented by a 4×1 unit length vector ψ shown in Eq. 5.

As mentioned above, states in quantum theory can be represented by a vector, such as ψ, or by what is called a density matrix. The density operator produced by the state vector ψ is formed from the outer product of the cognitive state: $\rho = \psi \cdot \psi^{\dagger}$. The diagonals of the density matrix contain the action probabilities.

The key new idea is that the density matrix evolves across time according to the following quantum master equation

$$\begin{aligned} \frac{d}{dt}\rho(t) &= -i \cdot (1-\alpha) \cdot [H, \rho] + \alpha \cdot \mathscr{L}(\rho), \\ \mathscr{L}(\rho) &= \sum \gamma_{ij} \cdot \left(L_{ij} \cdot \rho \cdot L_{ij}^{\dagger} - \frac{1}{2}\left\{ L_{ij}^{\dagger} \cdot L_{ij}, \rho \right\} \right), \\ [H, \rho] &= H \cdot \rho - \rho \cdot H, \\ \left\{ \left(L_{ij}^{\dagger} \cdot L_{ij} \right), \rho \right\} &= \left(L_{ij}^{\dagger} \cdot L_{ij} \right) \cdot \rho + \rho \cdot \left(L_{ij}^{\dagger} \cdot L_{ij} \right) \end{aligned} \tag{13}$$

where L_{ij} is a 4×4 matrix with a one located at cell (i, j) and zeros elsewhere. The master equation describes the dynamics of what is called an open quantum system.

The first part, $[H, \rho]$, of the sum in Eq. 13 represents von Neumann dynamics produced by a pure quantum system. The parameter, h_{ij}, in the 4×4 Hermitian matrix $H = H^{\dagger}$ determines the rate of change in transition to state i from another state j. For simplicity, they set

$$H = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

The second part, $\mathscr{L}(\rho)$, of the sum in Eq. 13 represents the Lindblad dynamics of a Markov system. The parameters $\gamma_{ij}, i, j = 1, \ldots, 4$ form a 4×4 Markov transition matrix, Γ,with γ_{ij} equal to the probability of transiting to state i from another state j. These parameters were determined from the payoffs to the player (e.g., Ann) depending on her anticipation regarding the opponent's (e.g., Bill's) move. At each moment in time, there is some probability ϕ that Ann switches from contemplating that Bill cooperates to defecting and visa versa. When contemplating that Bill cooperates, Ann considers defecting with a probability based on the payoff advantage for defecting when Bill cooperates; when contemplating that Bill defects, Ann considers defecting with a probability based on the payoff advantage for defecting when Bill defects.

Martínez-Martínez and Sánchez-Burillo (2016) applied the model to the disjunction effect as follows. When told Bill cooperated, Ann's initial state is set equal to $\rho_{11} = \rho_{22} = 1/2$, and $\rho_{33} = \rho_{44} = 0$; and it evolves for some period of time t using Eq. 13. During this evolution, the switching parameter ϕ fixed to zero so the Ann remains contemplating Bill to cooperate. Then at time t, using the projection matrix $P_C = diag\begin{bmatrix} 0 & 1 & 0 & 1 \end{bmatrix}$, the probability that Ann cooperates equals the trace: $p(C_A|C_B) = Tr\left[P_C \cdot \rho_C(t)\right]$. Likewise, when told Bill defected, Ann's initial state is set to equal $\rho_{33} = \rho_{44} = 1/2$, and $\rho_{11} = \rho_{22} = 0$. Again this state evolves with the switching probability fixed to zero so that Ann remains contemplating Bill to defect. The probability that Ann cooperates under this condition equals the trace: $p(C_A|D_B) = Tr\left[P_C \cdot \rho_D(t)\right]$. However, when Bill's move is unknown, the initial state remains equal to $\rho_{33} = \rho_{44} = 1/4$, and $\rho_{11} = \rho_{22} = 1/4$., and it evolves with the switching probability free so that Ann's attention can switch back and forth between Bill cooperating and Bill defecting. Then probability that Ann cooperates equals the trace: $p(C_A|U) = Tr\left[P_C \cdot \rho_U(t)\right]$.

This open system model provides a good way to combine quantum and Markov systems. The system starts out operating in the quantum regime to produce a coherent (quantum) density matrix with non-zero off-diagonal cells, but later the system ends in the Markov regime to produce a de-coherent (classical) density matrix with zero off-diagonal cells.

To account for the disjunction effect, Martínez-Martínez and Sánchez-Burillo (2016), set the initial conditions as follows. They used three free parameters: one for the probability to cooperate based on the payoffs, on for the switching probability, and the third was the parameter $0 \leq \alpha \leq 1$ that determines the weight of the contribution of each dynamic process. They were able to fit parameters that could very accurately reproduce the disjunction effect results: $p(C_A|C_B) = 0.16, p(C_A|D_B) = 0.09, p(C_A|U) = 0.34$ Also their analyses show that it is necessary for $(1 - \alpha) > 0$ to reproduce the disjunction effect (the quantum dynamics are essential). To fit the results of Shafir and Tversky (1992), their estimate of the weight equaled $\alpha = 0.81$. However, this model has not been developed to account for predictions that a player might make concerning an opponent's play, and consequently, it is unable to account for question order effects.

3.8 Tanaka Umegaki Nishiyama Kitoh-Nishioka model

Tanaka et al. (2022) proposed another type of dynamic open system model of the PD game. When applied to the disjunction effect, they also use a four dimensional state represented by Eq. 5, which they convert into a density matrix $\rho = \psi \cdot \psi^{\dagger}$. For the known and unknown conditions, they used the same initial conditions as used by Martínez-Martínez and Sánchez-Burillo (2016).

They assume that the state ρ evolves across time according to Eq. 13, but with $\alpha = 0$. Instead of using a Markov process, they assume that the cognitive state ψ is influenced

Table 1 Comparison of models with findings

Model	Finding Disjunction	Interference	Question order	Sequential	Promise
Tesar	N	Y	Y	N	?
Yukalov	Y	N	N	N	N
Pothos	Y	Y	Y	N	?
Denolf	N	?	N	Y	N
Kvam	?	?	?	N	Y
Asano	Y	N	N	N	N
Martinez	Y	N	N	N	N
Tanaka	Y	N	N	N	N

Y indicates the existing model can reproduce the finding, ? indicates that the existing model is capable but has not been applied to this finding, N indicates the existing model is not yet capable (but could be modified in the future) to reproduce the finding. Tesar refers to (Tesar, 2020), Yukalov refers to (Yukalov & Sornette, 2014), Pothos refers to (Pothos & Busemeyer, 2009), Denolf refers to (Denolf et al., 2016), Kvam refers to (Kvam et al., 2014), Asano refers to (Asano et al., 2011b), Martinez refers to (Martínez-Martínez & Sánchez-Burillo, 2016), Tanaka refers to (Tanaka et al., 2022)

by many unknown (noisy) environmental influences, which they directly model using Hamiltonians. The total Hamiltonian for Eq. 13 (with $\alpha = 0$) is composed of two parts: $H = H_S + H_E$, where H_S is the Hamiltonian for the system represented by ψ, and H_E is the environmental noise. The environment is modeled by a set of quantum-mechanical harmonic oscillators that produce noise with an amount controlled by a temperature parameter T. The environmental influences from H_E eventually stabilize the state toward an equilibrium state, and without the latter, the system would oscillate indefinitely, The Hamiltonian H_S is the key for producing the disjunction effect. They define the system Hamiltonian as

$$H_S = \begin{bmatrix} h_{DD} - \delta & \Delta & -\delta & 0 \\ \Delta & h_{DC} + \delta & 0 & -\delta \\ -\delta & 0 & h_{CD} + \delta & \Delta \\ 0 & -\delta & \Delta & h_{CC} - \delta \end{bmatrix} \tag{14}$$

where the diagonal elements are determined by the payoffs to Ann for each pair of actions. For example, h_{CD} is determined by the payoff Ann receives when Bill cooperates and she defects. The parameter δ is used to produce an effect related to what Shafir and Tversky (1992) called "wishful thinking." When the opponent's move is known, they set $\delta = 0$, and when the opponent's move is unknown they allow δ to be a free parameter to be estimate from the data. The value of parameter δ changes the payoffs of the game, that is, it increases the payoffs for both cooperating and both defecting and it decreases the payoffs for one player cooperating while the other player defects.

Using the two parameters, α, δ to determine the system Hamiltonian, H_S, and an additional parameter, T, for the environmental noise in H_E, they were able to approximately reproduce the disjunction effect: $p(C_A|C_B) = 0.22, p(C_A|D_B) = 0.22, p(C_A|U) = 0.44$. However, this model does provide any account for predictions that a player might make concerning an opponent's play, and consequently, it is unable to account for question order effects.

3.9 Summary of models

Table 1 provides a summary comparison of the eight models with the five findings. Of course, the responses in the table are based on the authors' judgments and the creators of the models might have a different opinion. At least the table reflects the finding that was the target of the each model. There are, of course, advantages and disadvantages of each model that are not reflected in the table. In particular, some of the models are designed to account for both predictions as well as actions (Tesar, 2020; Pothos & Busemeyer, 2009; Denolf et al., 2016; Kvam et al., 2014) while others have not developed this capability (Yukalov & Sornette, 2014; Asano et al., 2011b; Martínez-Martínez & Sánchez-Burillo, 2016; Tanaka et al., 2022). Some of the models (Pothos & Busemeyer, 2009; Asano et al., 2011b; Martínez-Martínez & Sánchez-Burillo, 2016; Tanaka et al., 2022) specify how the payoffs of the game map into the Hamiltonian used to predict the choice probabilities in the game, and some of the models don't explicitly describe any mapping (Tesar, 2020; Yukalov & Sornette, 2014; Denolf et al., 2016; Kvam et al., 2014). Some of the models (Asano et al., 2011b; Martínez-Martínez & Sánchez-Burillo, 2016; Tanaka et al., 2022) are specifically designed to describe the dynamics of the decision in the game. Another important issue concerns differences among the models with respect to complexity and number of parameters. Finally, some quantum models not reviewed here (Bagarello et al., 2017) have been applied to the PD game but not applied to the findings reviewed in this article. All of these models can of course be modified to address some of their limitations, and so this theory development is still under construction.

4 Discussion

In this article, we reviewed five different empirical findings obtained with the PD game: the disjunction effect, negative influence of predictions on actions in simultaneous play, positive influence of predictions on action in sequential play, and effects changing the order of predictions and action questions. These findings are all difficult for standard economic game theories to explain, but they all can be related to a violation of what is expected from the classical law of total probability. Quantum theory provides a natural explanation for violations of total probability based quantum interference effects produced by measurement. For this reason, a substantial amount of theoretical effort has been spent developing quantum models to explain these findings. However, these quantum models have never been systematically compared to all five findings and to each other. Furthermore, although these models have appeared in various fields ranging from psychology to physics, they are little known in the experimental economic field. The purpose of this article was to make these comparisons, which are summarized in Table 1, and bring them to the attention of experimental economists.

Future research is needed to empirically test new predictions of these models. One interesting prediction concerns what is called the A-B-A paradigm (a psychology version of the Stern–Gerlach experiment from physics). Suppose participants are asked to make a prediction at stage 1, followed by an action decision at stage 2, and then followed by a prediction again at stage 3, before observing the opponent's action at stage 4. Models in Table 1 that employ projectors to represent the measurements (i.e., the Tesar model and the Pothos model) must predict that the action decision at stage 2 changes the player's mind from stage 1 to stage 3 (Khrennikov et al., 2014). A more recent model of question order effects that employs quantum instruments rather than projectors to represent the measurements predicts no change from stage 1 to stage 3 (Ozawa & Khrennikov, 2019). This is important question that deserves further research.

One last comment concerns an interesting "interference" effect studied by (Rapoport et al., 2009) called the Braess Paradox. This refers to the counterintuitive observation that adding links to a directed transportation network with usage externalities may raise the costs of all users. I once discussed this finding with Amnon and suggested that perhaps there may be a quantum explanation for this paradox. He smiled with encouragement but I also detected a bit of skepticism in his eyes. This trace of doubt will be like a sharp sword, inspiring courage and motivation to cut open solutions to network paradoxes.

Appendix A: Bounds on interference

This appendix derives the bound on the Yukalov and Sornette (2011) model for the PD game described in Subsect. 3.2. According to their quantum model, the interference is derived to be

$$\begin{aligned}
2 \cdot Real\left(\phi_1^* \cdot \psi_{CC} \cdot \phi_2 \cdot \psi_{DC}^*\right) &= 2 \cdot \left|\phi_1^* \cdot \psi_{CC} \cdot \phi_2 \cdot \psi_{DC}^*\right| Real(cos(\theta) + i \cdot sin(\theta)) \\
&= 2 \cdot \left|\phi_1^* \cdot \psi_{CC} \cdot \phi_2 \cdot \psi_{DC}^*\right| \cdot cos(\theta) \\
&= 2 \cdot \left|\phi_1^* \cdot \psi_{CC} \cdot \phi_2 \cdot \psi_{DC}^*\right| \cdot cos(\theta) \\
&= 2 \cdot \left|\phi_1^* \cdot \psi_{CC}\right| \cdot \left|\phi_2 \cdot \psi_{DC}^*\right| \cdot cos(\theta) \\
&= 2 \cdot \sqrt{p_{CC}} \cdot \sqrt{p_{DC}} \cdot cos(\theta)
\end{aligned}$$

Applying the data from Shafir and Tversky (1992) we obtain

$$\begin{aligned}
2 \cdot \sqrt{p_{CC}} \cdot \sqrt{p_{DC}} \cdot cos(\theta) &= 2 \cdot \left(\sqrt{0.5 \cdot 0.16}\right)\left(\sqrt{0.5 \cdot 0.03}\right) \cdot cos(\theta) \\
&= 0.0693 \cdot cos(\theta) < +0.25
\end{aligned}$$

Applying the data from Busemeyer and Matthews (2006)

$$2 \cdot \sqrt{p_{CC}} \cdot \sqrt{p_{DC}} \cdot cos(\theta) = 2 \cdot \left(\sqrt{0.5 \cdot 0.16}\right)\left(\sqrt{0.5 \cdot 0.09}\right) \cdot cos(\theta)$$
$$= 0.12 \cdot cos(\theta) < +0.25$$

Acknowledgements This research was supported by AFOSR Grant number FA9550-20-1-0027 to the first author.

Funding Air Force Office of Scientific Research (FA9550-20-1-0027).

Declarations

Conflict of interest All authors have no financial interest.

References

Alonso-Sanz, R. (2019). *Quantum game simulation* (Vol. 36). Springer.

Asano, M., Basieva, I., Khrennikov, A., Ohya, M., & Tanaka, Y. (2012). Quantum-like dynamics of decision-making. *Physica A: Statistical Mechanics and its Applications, 391*(5), 2083–2099.

Asano, M., Ohya, M., & Khrennikov, A. (2011). Quantum-like model for decision making process in two players game. *Foundations of Physics, 41*(3), 538–548.

Asano, M., Ohya, M., Tanaka, Y., Basieva, I., & Khrennikov, A. (2011). Quantum-like model of brain's functioning: Decision making from decoherence. *Journal of Theoretical Biology, 281*(1), 56–64.

Bagarello, F., Haven, E., & Khrennikov, A. (2017). A model of adaptive decision-making from representation of information environment by quantum fields. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375*(2106), 20170162.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2014). Preferences and beliefs in a sequential social dilemma: A within-subjects analysis. *Games and Economic Behavior, 87*, 122–135.

Broekaert, J. B., Busemeyer, J. R., & Pothos, E. M. (2020). The disjunction effect in two-stage simulated gambles. An experimental study and comparison of a heuristic logistic, markov and quantum-like model. *Cognitive Psychology, 117*, 101262.

Bruza, P. D., Fell, L., Hoyte, P., Dehdashti, S., Obeid, A., Gibson, A., & Moreira, C. (2023). Contextuality and context-sensitivity in probabilistic models of cognition. *Cognitive Psychology, 140*, 101529.

Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambirdge University Press.

Busemeyer, J. R., & Lambert-Mogiliansky, A. (2009). Empirical comparison of Markov and quantum models of decision making. *Journal of Mathematical Psychology, 53*(5), 423–433.

Busemeyer, J. R. , & Matthews, M. (2006). A quantum information processing explanation of disjunction effects. In R. Sun & N. Myake (Eds.), Proceedings of the 29th annual conference of the cognitive science society and the 5th international conference of cognitive science (pp. 131–135). Erlbaum.

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

Denolf, J., Martínez-Martínez, I., Josephy, H., & Barque-Duran, A. (2016). A quantum-like model for complementarity of preferences and beliefs in dilemma games. *Journal of Mathematical Psychology, 78*, 96–106.

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

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.

Gleason, A. M. (1957). Measures on the closed subspaces of a hilbert space. *Journal of Mathematical Mechanics, 6*, 885–893.
Hughes, R. I. G. (1989). *The structure and interpretation of quantum mechanics*. Harvard University Press.
Khrennikov, A. Y. (2010). *Ubiquitous quantum structure: From psychology to finance*. Springer.
Khrennikov, A. Y., Basieva, I., Dzhafarov, E. N., & Busemeyer, J. R. (2014). Quantum models for psychological measurements: An unsolved problem. *PLoS ONE, 9*(10), e110909.
Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. Chelsea Publishing Co.
Kvam, P. D., Busemeyer, J. R., Lambert-Mogiliansky, A. (2014). An empirical test of type indeterminancy in the prisoners-dilemma. In B. H. Atmanspacher, E. Haven, K. Kitto, & D. Rene (Eds.), Quantum interaction, 7th international symposium (pp. 213–224). Springer.
Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence. *Proceedings of the National Academy of Science, 112*(34), 10645–10650.
Lambert-Mogiliansky, A., Zamir, S., & Zwirn, H. (2009). Type indeterminacy: A model of the “kt” (Kahneman-Tversky)-man. *Journal of Mathematical Psychology, 53*(5), 349–361.
Martínez-Martínez, I., & Sánchez-Burillo, E. (2016). Quantum stochastic walks on networks for decision-making. *Scientific Reports, 6*, 23812.
Meyer, D. A. (1999). Quantum strategies. *Physical Review Letters, 82*, 1052–1055.
Ozawa, M., & Khrennikov, A. (2019). Application of theory of quantum instruments to psychology: Combination of question order effect with response replicability effect. *Entropy, 22*(1), 37.
Piotrowski, E. W., & Sladkowski, J. (2002). An invitation to quantum game theory. *International Journal of Theoretical Physics, 42*, 1089–1099.
Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability model explanation for violations of ‘rational’ decision making. *Proceedings of the Royal Society B, 276*(1665), 2171–2178.
Pothos, E. M., & Busemeyer, J. R. (2022). Quantum cognition. *Annual Review of Psychology, 73*, 749–778.
Rapoport, A. (1988). Experiments with n-person social traps i: Prisoner’s dilemma, weak prisoner’s dilemma, volunteer’s dilemma, and largest number. *Journal of Conflict Resolution, 32*(3), 457–472.
Rapoport, A., Kugler, T., Dugar, S., & Gisches, E. J. (2009). Choice of routes in congested traffic networks: Experimental tests of the Braess paradox. *Games and Economic Behavior, 65*(2), 538–571.
Rapoport, A., & Wallsten, T. S. (1972). Individual decision behavior. *Annual Review of Psychology, 23*(1), 131–176.
Santos, A. C. (2020). Entanglement and coherence in quantum prisoners dilemma. *Quantum Information Processing, 19*(1), 13.
Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology, 24*, 449–474.
Tanaka, S., Umegaki, T., Nishiyama, A., & Kitoh-Nishioka, H. (2022). Dynamical free energy based model for quantum decision making. *Physica A: Statistical Mechanics and its Applications, 605*, 127979.
Tesar, J. (2020). A quantum model of strategic decision-making explains the disjunction effect in the prisoner’s dilemma game. *Decision, 7*(1), 43–54.
Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science, 3*, 305–309.
Von Neumann, J. (1955). *Mathematical foundations of quantum theory*. Princeton University Press.
White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). Sometimes it does hurt to ask: The constructive role of articulating impressions. *Cognition, 1*, 48–64.
Yukalov, V. I., & Sornette, D. (2011). Decision theory with prospect interference and entanglement. *Theory and Decision, 70*, 283–328.
Yukalov, V. I., & Sornette, D. (2014). How brains make decisions. *Universe of scales: From nanotechnology to cosmology* (pp. 37–53). Springer.