City Research Online

City St George's, University of London

Citation: Busemeyer, J. R. & Pothos, E. M. (2012). Social Projection and a Quantum Approach for Behavior in Prisoner's Dilemma. Psychological Inquiry, 23(1), pp. 28-34. doi: 10.1080/1047840x.2012.652488

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/4714/

Link to published version: https://doi.org/10.1080/1047840x.2012.652488

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Social projection and a quantum approach for behavior in Prisoner's Dilemma

Jerome R. Busemeyer[1] and Emmanuel M. Pothos[2]

1. Department of Psychology, Indiana University, 10th St., Bloomington, IN 47405, USA. Email: *jbusemey@indiana.edu*.

2. Department of Psychology, Swansea University, Swansea SA2 8PP, UK. Email: *e.m.pothos@swansea.ac.uk* .

In their target article, Krueger, DiDonato, and Freestone provided a detailed and thought-provoking critique of existing approaches to understanding human behavior in social dilemmas, including the famous Prisoner's Dilemma paradigm. In this way, they motivate their own approach, based on the idea of social projection. This commentary establishes some theoretical relations between the social projection hypothesis and a new "quantum inspired" model of behavior for the prisoner's dilemma game (Pothos & Busemeyer, 2009). Our discussion reveals a subtle difference in the motivation between the social projection hypothesis and the "quantum inspired" model, a commonality of assumptions (despite the obviously different form of the models), and a complementarity of objectives (and strengths). Further elaboration of the two models may well lead to a convergence between them, which will further enhance both our insight of the factors which drive human intuition in social dilemmas and of the relevant formal principles.

One of the most famous paradigms for studying social dilemmas and human cooperative behavior is Prisoner's Dilemma (PD), in which a player can decide to cooperate (C) or defect (D) with another (imaginary or otherwise) player, who can also C or D. A payoff matrix determines the reward of the player depending on her choice and on the actions of the other player. The payoff matrix is typically set up in such a way that if both players C they receive a relatively high amount (say $20) and if they both D a relatively low amount (say $10). But, if the first player chooses to D and the second chooses to C, then the first player receives the highest possible reward ($25) and the second player the lowest possible reward ($5; and vice versa). Perhaps the PD paradigm has attracted the attention it has because it embodies one of the fundamental riddles of why human societies work: mutual cooperation leads to success (prosperity) for all members, but individuals who choose to D stand to gain the most (as long as the majority C).

A classical rationality view simply predicts that each player should D (the Nash equilibrium is for both players to D). In fact, in laboratory versions of the PD task, naïve observers often choose to C and so ignore the prescription from classical (selfish) rationality. Understanding the psychological mechanism which makes a C decision more favorable has proved a major theoretical challenge, and

is the focus of the target article of Krueger, DiDonato, and Freestone ,titled “Social Projection Can Solve Social Dilemmas” (henceforth we will refer to it as just the ‘target article’ and the authors as KDF).

KDF provide a detailed overview of alternative theories to account for cooperative behavior in social dilemmas, such as PD. For example, they discuss a morality theory, according to which cooperation is favored because it is considered a more just or fair strategy. A reciprocity theory involves assumptions regarding expectations of how the other player ought to behave. A benevolence (social values) hypothesis computes expected reward for a player as a weighted sum of the player reward and the reward of the other player. Team reasoning approaches consider reward from the point of view of the entire team, not just the individual players. An error approach assumes that the optimal strategy is indeed to D, however, players simply occasionally makes errors. Yet another explanation for cooperative behavior has to do with observations (or assumptions) regarding the preponderance of CC situations (that is, situations in which both players decide to C) in particular types of games.

KDF’s critical evaluation of these approaches reveals several weaknesses and so motivates their own explanation. This depends on social projection, that is, the idea that naïve observers often believe that others will behave as they do. Social projection allows a naïve observer to calculate an estimate of expected gain for cooperating, under particular assumptions of correspondence in the actions of the two players. KDF support their hypothesis with a series of experiments, which are based on variations of the PD task. Their social projection hypothesis is the one most consistent with the empirical findings, in terms of predicting participant choices (specifically, the recommendations for C or D from a participant towards an imaginary player whose actions are given), of predicting judgments of rationality for the imaginary player, and of predicting judgments of morality for the imaginary player. Support for the KDF is thoughtful, innovative, and, ultimately, convincing.

The purpose of this commentary is to consider Pothos and Busemeyer’s (2009) modeling of PD (and two-stage gambling task) data, on the basis of quantum probability (henceforth, QP). Given

our optimistic view of KDF's theorizing, one can wonder as to what could be gained by introducing yet another theory and, indeed, one which is based on unfamiliar formal principles (why is QP needed?). As we shall discuss, first, Pothos and Busemeyer's (2009) quantum approach enables an additional, fundamental perspective regarding the rationality of behavior in a PD task. Second, while KDF show that individually any of the existing approaches fail, the quantum theory provides a formalization of a *combination* scheme, which appears to work. In fact, the principles of this combination scheme align well with those of the social projection theory. Ultimately, the objectives of the quantum model are somewhat different from those of KDF. For KDF, the key objective is to explain why naïve observers find it attractive to C in particular PD situations. For the quantum model, the challenge is to reconcile behavior in PD situations with formal probability theory. The importance of this second issue and the related fundamental implications for the rationality debate are considered next.

A different perspective on the rationality of the PD task

Shafir and Tversky (1992; Tversky & Shafir, 1992) discussed a variation of the PD task in which, in some trials, players were simply told of the opponent's decision. Unsurprisingly, when participants were told that the other person chose to D, they chose do D: Prob(D|D)=.97. When participants were told that the other person decided to C, they also chose to D, so that Prob(D|C)=.84 (a one-shot PD task was employed, so there were no grounds to worry about tit for tat or long term strategies). But, when participants were not told about the other player's action, the probability to C rose, and Shafir and Tversky (1992) observed a Prob(D|unknown)=.63 (for replications see Crosson, 1999; Li & Taplan, 2002; Busemeyer, Matthew, & Wang, 2005). Similar results were obtained with a two-stage gambling task paradigm, in which participants had to decide whether or not to play a second gamble on the basis of knowledge (or not) of the outcome of a first gamble (Tversky & Shafir, 1992).

Several researchers advocate the view that human naïve decision making and judgment follows the principles of classical (Bayesian) probability (CP) theory. Without doubt, there are some

aspects of cognitive process which can be successfully described with CP principles (e.g., Anderson, 1991; Griffiths et al., 2010; Tenenbaum et al., 2011). Indeed, some researchers have suggested that human cognition *has* to follow the principles of CP theory, because it is rational to do so (e.g., Oaksford & Chater, 2007; cf. Pothos & Busemeyer, in press). For example, the so-called Dutch-book theorem guarantees that, as long as you assign probabilities to bets on the basis of CP theory, you will neither lose nor gain money.

Unfortunately, for all its normative support, CP theory fails to explain Shafir and Tversky's (1992) results. One of the fundamental axioms of CP theory is the law of total probability, which can be expressed as

$$P(A) = P(A \cup (X \cap \bar{X})) = P(A \cap X) + P(A \cap \bar{X}) = P(X)P(A|X) + P(\bar{X})P(A|\bar{X}).$$

In other words, the probability of an event *A* is the sum of the probability of the conjunction of *A* with another event *X* and the conjunction of *A* with *X*'s negation. The law of total probability is intuitive: The opponent can either D with some probability $P(D_{Player\ 2})$ or C with some probability $P(C_{Player\ 2})$. If the opponent is predicted to D, then the player has some probability for choosing D, $P\left(D_{Player\ 1}|D_{Player\ 2}\right)$; or if the opponent is predicted to C, then the player has some other probability for choosing D, $P\left(D_{Player\ 1}|C_{Player\ 2}\right)$. Therefore, when the opponent's action is unknown, the probability of defection should equal

$$P\left(D_{Player\ 1}\right) = P(D_{Player\ 2})P\left(D_{Player\ 1}|D_{Player\ 2}\right) + P(C_{Player\ 2})P\left(D|C_{Player\ 2}\right).$$

Note that $P\left(D_{Player\ 1}\right)$ is bounded by the two conditional probabilities, $P\left(D_{Player\ 1}|D_{Player\ 2}\right)$ and $P\left(D_{Player\ 1}|C_{Player\ 2}\right)$, so that it cannot be lower than the lowest of these two conditional probabilities or higher than the highest of these conditional probabilities. Empirically, however, the observed probabilities do not obey these bounds in the case of Shafir and Tversky's (1992) variation to the PD game. In fact Shafir and Tversky (1992) found that $P\left(D_{Player\ 1}\right)$ was lower than both conditional probabilities. Whichever way you manipulate the prior assumptions for whether the opponent is likely to C or D, Shafir and Tversky's (1992) findings cannot be reconciled with the law of total probability.

So, if it is rational for cognitive process to be consistent with the principles of CP theory, and if naive observers violate these principles in the case of, e.g., Shafir and Tversky's (1992) PD task, do we have to conclude that people are (mostly) irrational? According to Shafir, Tversky, Kahneman, and their colleagues, either we need to accept this bleak conclusion for human irrationality, or adopt their proposal that CP theory has little to do with understanding cognitive processes. In a research program that has had an enormous impact in psychology (e.g., over 30,000 citations to Tversky's work and a Nobel prize for Kahneman in 2002), Shafir, Tversky, Kahneman, and their colleagues reported several violations of CP theory in naïve decision making (e.g., Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1973, 1974).

The first point of this commentary is that none of the theories KDF reviewed can readily offer some reconciliation between the violation of the law of total probability and the view of human rationality based on adherence to CP principles. Note, again, that the rationality we have discussed in this section is different from the one KDF examine, in relation to their model. KDF are concerned with whether participant choices to D or C can be considered as rational, given expectations for reward. We presented a more general notion of rationality, which arises from a consideration of PD behavior from the perspective of formal probability theory. We do not claim that any of the accounts presented in KDF's paper are necessarily inconsistent with a resolution of the conflict between the violation of the law of total probability and Shafir and Tversky's (1992) empirical findings. Rather, a possible resolution is not discussed.

Quantum probability theory

For most psychologists, when it comes to the question of how to systematically assign probabilities to events, CP theory is all there is. CP theory is so ingrained that it is hard to even imagine alternative probabilistic frameworks. For example, how can it not be that $P(A) = P(A \cap X) + P(A \cap \bar{X})$? Nevertheless, in the beginning of the 20th century, physicists realized that the way CP theory formalizes uncertainty is often inconsistent with physical observation. As a result, and over a period of around three decades, physicists developed quantum mechanics, arguably one

of the most sophisticated creations of the human mind, and one which have had an unparalleled impact on human history (through the development of transistors, the foundations of chemistry, the development of lasers, etc.).

Quantum mechanics is a theory of physics, which is based on a particular theory of probability, QP theory. QP theory can be dissociated from its physical origins and is potentially applicable in any area of human inquiry in which there is a need to formalize uncertainty. For example, QP theory has been applied in economics (e.g., Baaquie, 2004) and information theory (e.g., Grover, 1997). Moreover, the potential relevance of QP theory in psychology can be motivated a priori, in terms of various key characteristics of QP theory which appear consistent with general intuition about cognitive process (Busemeyer and Bruza, 2012).

The first characteristic concerns the quantum concept of superposition. Classic cognitive models assume that at each moment a person is in a definite state with respect to some judgment. Of course, it is not known what the person's true state is at each moment, and so the model can only assign a probability to a response with some value at each moment. But the model is stochastic only because it does not know exactly what trajectory (definite state at each time point) a person is following. In this sense, cognitive scientists currently model the cognitive system as if it were a particle producing a definite sample path through a state space. Quantum theory works differently by allowing a person to be in an indefinite state (formally called a superposition state) at each moment in time. Strictly speaking, being in an indefinite or superposition state means that the model cannot assume that you have a definite value, with respect to some judgment scale, at each moment in time. You can be in an indefinite state that allows all of these definite states to have potential. A superposition state perhaps provides a better representation of the conflict, ambiguity, or uncertainty that people experience at each moment. In this sense, quantum theory allows one to model the cognitive system as if it were a wave moving across time over the state space. This would be so until a decision is made, at which point the state is forced to change from an indefinite state to a definite response.

A second reason concerns a putative sensitivity of the cognitive system to measurement. Traditional cognitive models assume that whatever we record at a particular moment reflects the state of the system, as it existed immediately before we inquired about it. The answer to a judgment question simply reflects the state regarding this question just before we asked it. One of the more provocative lessons learned from quantum theory is that taking a measurement of a system creates, rather than records, a property of the system. Immediately before asking a question, a quantum system can be in an indefinite state. The answer we obtain from a quantum system is constructed from the interaction of the indefinite state and the question that we ask. This interaction creates a definite state out of an indefinite state. We argue that the quantum principle, that a reality is constructed from an interaction between the person's indefinite state and the question being asked, better matches psychological intuition for complex judgments, than the assumption that an answer simply reflects a pre-existing state.

The third reason concerns the quantum concept of measurement incompatibility. The change in state that results after answering one question causes a person to respond differently to subsequent questions. Answering one question disturbs the answers to subsequent questions and the order of questioning becomes important. In other words, the first question sets up a context which changes the answer to the next question. Consequently, we cannot define a joint probability of simultaneous answers to a conjunction of questions, and instead we can only assign a probability to the sequence of answers. In quantum physics, order dependent measurements are said to be non-commutative and quantum theory was especially designed for these types of measures. Many of the mathematical properties of quantum theory arise from developing a probabilistic model for non-commutative measurements, including Heisenberg's (1927) famous uncertainty principle. Question order effects (e.g., Moore, 2002) are major concern for attitude researchers, and a theoretical understanding of such effects can be achieved within quantum theory.

The fourth reason is that human judgments do not always obey classic laws of logic and probability. The classic probability theory used in current cognitive and decision models is derived

from the Kolmogorov axioms, which assign probabilities to events defined as sets. Consequently, the family of sets in the Kolmogorov theory obey the Boolean axioms of logic, and one important axiom of Boolean logic is the distributive axiom. From this distributive axiom, one can derive the law of total probability, which provides the foundation for inferences with Bayes nets. However, the law of total probability is violated by the results of many psychological experiments, including Shafir and Tversky's (1992) variation of the PD task. Quantum probability theory is derived from the Dirac and von Neumann axioms. These axioms assign probabilities to events defined as subspaces of a vector space, and the logic of subspaces does not obey the distributive axiom of Boolean logic. The fact that quantum logic does not always obey the distributive axiom implies that the quantum model does not always obey the law of total probability. Essentially, quantum logic is a generalization of classic logic and quantum probability is a generalized probability theory. Classic probability theory may be too restrictive to explain human judgments.

Although the use of quantum principles for modeling psychological processes is still in its infancy, there have already been some very promising results (e.g., Aerts & Gabora, 2005; Atmanspacher, Filk, & Romer, 2004; Blutner, 2009; Bordley, 1998; Bruza, Kitto, Nelson, & McEvoy, 2009; Busemeyer, Wang, & Townsend, 2006; Busemeyer et al., 2011; Khrennikov, 2010; Lambert-Mogiliansky, Zamir, & Zwirn, 2009; Pothos & Busemeyer, 2009; Yukalov & Sornette, 2010). Now that we have identified some general reasons for considering a quantum approach to cognition and decision, we present our quantum inspired model for the PD game and relate these ideas to the social projection hypothesis.

Quantum model of social projection

Pothos and Busemeyer (2009) formulated a quantum model for the PD game , which incorporates the idea of social projection, according to which a player's particular action implies a belief in a corresponding action by the opponent (in our original work we described this influence in terms of cognitive dissonance theory, Festinger, 1957, and wishful thinking, but a characterization as

social projection is more accurate). The main ideas are briefly sketched below (see Pothos & Busemeyer, 2009, for the full mathematical details).

The decision maker's tendencies to make inferences and take actions are based on a state vector, which is a unit length vector lying within a four dimensional space. The four coordinates of the state represent the "probability amplitudes" for the four events DD, DC, CD, and CC, where the first letter indicates the prediction about the opponent's action and the second letter indicates the player's action. For example, CD represents the event that the opponent is predicted to C but the player decides to D.

The player begins the process with an *initial* state, denoted by ψ, that contains information provided about the opponent's action. This initial state concerns the player's initial biases for action based on corresponding beliefs about the opponent's intentions. The initial state can be affected by information the player receives about the opponent's intentions. If the player is informed that the opponent will D, then the probability amplitudes for DC and DD are initially equal to zero, to produce a unit length state denoted ψ_D; if the player is informed that the opponent will C, then the probability amplitudes for CC and CD are initially equal to zero, to produce a unit length state denoted ψ_C; and if the opponent's action is unknown, then the player's initial state is a superposition of the two known states

$$\psi_U = d \cdot \psi_D + c \cdot \psi_C \ , \ |d|^2 + |c|^2 = 1.$$

This superposition state is interpreted as follows. The player does not necessarily predict that the opponent will D, and at the same time, the agent does not necessarily predict that the opponent will C. Also the player does not predict both contradictory events to occur at the same time. Instead the player is in an *indefinite* state, in which there is some potential for each prediction to be made.

The initial state ψ is then transformed into a final state φ. The transformation represents the thought process, during which the player clarifies her beliefs about the opponent's actions and her own decisions. During this deliberation period, the state is "rotated" by a unitary matrix, denoted $\boldsymbol{U}$, to produce a revised state $\varphi = \boldsymbol{U} \cdot \psi$, which remains unit length. The unitary matrix "rotates" the state

in a direction that is determined by two factors. One factor is the utility of the payoffs, which generates a potential for defection in the PD game (increasing the probability amplitudes for CD and DD). The second factor corresponds to KDF's social projection influence, which generates a tendency for beliefs and actions to become aligned (increasing the probability amplitudes for CC and DD). Both factors work together in a dynamic manner to change the initial state into a final state, as a result of the deliberation process.

Specifically, if the opponent's action is known D, then we have $\psi = \psi_D$, so that, $\boldsymbol{U}\cdot\psi_D = \varphi_D$, if the opponent's action is known C, then we have $\psi = \psi_C$, so that, $\boldsymbol{U}\cdot\psi_C = \varphi_C$, and if the opponent's action is unknown, then $\boldsymbol{U}\cdot\psi_U = \boldsymbol{U}\cdot(d\cdot\psi_D + c\cdot\psi_C) = d\cdot\boldsymbol{U}\cdot\psi_D + c\cdot\boldsymbol{U}\cdot\psi_C = d\cdot\varphi_D + c\cdot\varphi_C$. The final state $\varphi = \boldsymbol{U}\cdot\psi$ that results after deliberation is another unit length vector, lying within the four dimensional space, which assigns new probability amplitudes to the four events DD, DC, CD, CC. The player's decision to D depends on the two probability amplitudes in the state φ assigned to DD and CD; the decision to C depends on the two probability amplitudes in the state φ assigned to DC and CC. For this reason, it is convenient to decompose the state into two orthogonal parts, depending on whether the player decides to D or C. Let us denote as α the projection on the defection action for the player and as β the projection on the cooperation action for the player, indexed in such a way so as to indicate knowledge of the opponent's actions. Then, $\varphi_D = \alpha_D + \beta_D$, so that P(player D|known D) $= ||\alpha_D||^2$ and $\varphi_C = \alpha_C + \beta_C$, so that P(player D|known C) $= ||\alpha_C||^2$.

The key issue concerns that happens in the case when the opponent's action is unknown. Using the above equations, we have:

$$\boldsymbol{U}\cdot\psi_U = \boldsymbol{U}\cdot(d\cdot\psi_D + c\cdot\psi_C) = d\cdot\boldsymbol{U}\cdot\psi_D + c\cdot\boldsymbol{U}\cdot\psi_C = d\cdot\varphi_D + c\cdot\varphi_C = d\cdot(\alpha_D + \beta_D) + c\cdot(\alpha_C + \beta_C),$$

The part of the state vector corresponding to the possibility that the player decides to D is given by $d\cdot\alpha_D + c\cdot\alpha_C$. Therefore, the probability that the player decides to D is:

$$||\, d\cdot\alpha_D + c\cdot\alpha_C\,||^2 = |d|^2\cdot||\,\alpha_D||^2 + |c|^2\cdot||\,\alpha_C||^2 + |d|\cdot|c|\cdot(\alpha_D{}'\alpha_C) + |c|\cdot|d|\cdot(\alpha_C{}'\alpha_D).$$

In other words, the probability of defection for the unknown condition equals

P(player D|unknown) = $|d|^2 \cdot$P(player D|known D) + $|c|^2 \cdot$P(player D|known C) + *Int*,

where the interference term equals $Int = (d^* \cdot c) \cdot (\alpha_D' \alpha_C) + (c^* \cdot d) \cdot (\alpha_C' \alpha_D)$. The interference term may be positive, negative, or zero, depending on the angle between the pair of vectors (α_C , α_D). If the interference term is zero, then QP agrees exactly with CP theory, and obeys the law of total probability. However, to account for the empirical findings, the interference term needs to be negative. This is where the social projection factor that affects the rotation of the final state becomes important. The social projection factor can make this angle negative, so that the probability of defection for the unknown condition falls below both of the probabilities for the known states.

Thus, time evolution during payoff evaluation in the quantum model leads to interference effects and such interference effects allow coverage of the empirical results in the basic PD paradigm. Simply put, probabilities in the quantum model are computed through a squaring operation (of amplitudes), so that $|a + b|^2 = a^2 + b^2 + a^* b + b^* a$. The first two terms are the classical terms, the last two terms the interference terms. In the quantum model, two individually good reasons for performing an action (e.g., defecting, in a PD task) can cancel each other out, when they are both present. For example, in the PD task, the (good) reason for defecting when the opponent defects and the (equally good) reason for defecting when the opponent cooperates, are not consistent with these other. So, when they are both present, the model generates these interference effects, and the law of total probability can be violated. In fact, the quantum model provides a rigorous formalization of Shafir and Tversky's (1992) suggestion that participants behave in the way they do in PD tasks (and related paradigms) because of a failure of consequential reasoning.

One can wonder whether the approach of combining a factor depending on the utility of payoffs and a factor depending on social projection could work within a classical (Markov) framework for computing probabilities. In brief, this is not possible (see Pothos & Busemeyer, 2009, for further details). In the classical model, the state vector $\psi_U = d \cdot \psi_D + c \cdot \psi_C$ would be a

superposition of probabilities. Therefore, $\boldsymbol{U}\cdot\psi_U = \boldsymbol{U}\cdot(d\cdot\psi_D + c\cdot\psi_C) = d\cdot\boldsymbol{U}\cdot\psi_D + c\cdot\boldsymbol{U}\cdot\psi_C = d\cdot\varphi_D + c\cdot\varphi_C = d\cdot(\alpha_D + \beta_D) + c\cdot(\alpha_C + \beta_C)$ would likewise directly involve probabilities, so that,

P(player D|unknown) = d·P(player D|known D) + c·P(player D|known C)

Note that the classical ψ, φ, and $\boldsymbol{U}$ can be specified in a way analogous to how they were specified in the quantum model, but of course the classical and quantum components are not identical. Crucially, the classical expression for P(player D|unknown) is analogous to the one derived for the quantum model, but without the interference term. This is another way to say that a classical model is always constrained by the law of total probability, regardless of the form of time evolution which is employed.

Going back to our discussion of KDF's proposal, it is clear that the quantum approach is based on some of the same ideas as KDF's social projection hypothesis. However, *just* the ideas of classical rationality and social projection are not sufficient to account for the correct pattern of results in PD. As we have seen, a CP model based on exactly these principles is unable to capture violations of the sure thing principle. By contrast, the QP model can capture such violations, because of interference terms which can arise in probabilistic computation. It is also worth pointing out that some aspects of the quantum model can be seen as specific computational formalizations of some aspects of the social projection model, notably the idea of social projection. The possible relation between the two approaches is further considered in the next section.

Future directions and final thoughts

An important aspect of KDF's demonstration is that they do not restrict themselves to predicting that many participants choose to C (or, in their paradigm, to recommend cooperation), but rather they provide empirically accurate predictions across a range of variations of the main PD paradigm. Can the quantum model for the PD task reproduce these predictions as well? This is an interesting question, but one that warrants considerable additional work. Although the principles and form of the quantum model are straightforward, its dynamics are complex. The parameters of the model interact and, moreover, can affect not just the amplitude, but also the period of

oscillation as well. Therefore, before the model can be extended to provide a more complete coverage of results in the PD task and, notably, the insightful variations KDF developed, a more complete theory of what is the appropriate time point to extract probabilities from the state vector is required. Nonetheless, predicting a violation of the law of total probability from a formal probabilistic model is already quite a challenge and this makes us optimistic that these further challenges can be overcome as well.

This leads naturally to the question of whether a quantum approach provides a better account of the (baseline) pattern of results in the PD task only because it affords more computational flexibility, compared to the classical one. After all, while the quantum PD model can violate the law of total probability, without the interference terms it can also be seen to obey the law of total probability. Therefore, is it the case that the quantum model is all of the classical model, so to say, and a little bit more (with the introduction of the interference terms)?

This is a complicated issue (cf. Myung, 2000; Pitt et al., 2002) but, in brief, there are plenty of indications that quantum models are not, in general (just) more flexible than classical ones. First, the quantum model is quite general and it has been shown to explain not only the violations of total probability found with the PD game (Pothos & Busemeyer, 2009), but also other violations of CP theory including conjunction and disjunction fallacies (Busemeyer, et al., 2011) and order effects on inference (Trueblood and Busemeyer, 2011). Second, even though quantum models can violate certain key constraints of classical models, notably the law of total probability, they are subject to alternative, highly restrictive constraints. For example, the matrices (like $\boldsymbol{U}$, above), which determine the dynamics of a quantum model, have to obey a property called double stochasticity. In general, CP and QP theory are founded from different sets of axioms (the Kolmogorov and Dirac/ von Neumann axioms respectively). Ultimately, whether the success of a model over an alternative one is due to flexibility is a technical issue which can only be assessed in the context of specific models. So far, this has only been done once. Shiffrin and Busemeyer (2011) compared a quantum model and a matched classical one in terms of both model fit and complexity, using a Bayesian approach.

Overall, across two different assumptions for the prior distribution of parameters, the quantum model was shown to be superior to the classical one. So, while the question of relative flexibility between the quantum and the classical models is clearly still open, the existing results are favoring the quantum models.

Overall, our discussion converges to two broad conclusions. First, the KDF model and the quantum model for behavior on the PD task have somewhat different objectives and this endows them with complementary strengths. The KDF model was built on the basis of a thorough and comprehensive consideration of the relevant psychological processes which could influence decisions in PD tasks. By contrast, the starting point of the quantum model was the interest in the more general issue of whether it is possible to reconcile human behavior in PD tasks with violations of the law of total probability; as discussed, this question is at the heart of the debate of whether it is desirable (many cognitive scientists think it is) and possible (researchers following Tversky and Kahneman's tradition think it is not) to attempt to understand cognition within a formal probabilistic framework. Second, the KDF model and the quantum model, as currently specified, appear fairly consistent with each other. It is possible that further psychological elaboration of the quantum model and further computational elaboration of the KDF model may lead to inconsistencies between the two approaches. Equally, it is possible that such an endeavor will converge the two approaches to the same, or very similar, models. We are looking forward to this challenge.

References

Aerts, D., & Gabora, L. (2005). A theory of concepts and their combinations II: A Hilbert space representation. Kybernetes, 34, 192-221.

Atmanspacher, H., & Filk, T. (2010). A proposed test of temporal nonlocality in bistable perception. Journal of Mathematical Psychology, 54, 314-321.

Anderson, J. R. (1991). The Adaptive Nature of Human Categorization. Psychological Review, 98, 409-429.

Baaquie, B. E. (2004) Quantum finance: Path integrals and Hamiltonians for options and interest rates. Cambridge University Press.

Blutner, R. (2009). Concepts and bounded rationality: An application of Niestegge's approach to conditional quantum probabilities. In L. e. a. Acardi (Ed.), Foundations of probability and physics-5 (Vol. 1101, p. 302- 310).

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

Bruza, P. D., Kitto, K., Nelson, D. & McEvoy, C. L. (2009). Is there something quantum-like about the human mental lexicon? Journal of Mathematical Psychology, vol. 53, pp. 362-377.

Busemeyer, J. R. & Bruza, P. (2011). Quantum models of cognition and decision making. Cambridge University Press: Cambridge, UK.

Busemeyer, J. R. , Matthew, M., & Wang, Z. A.(2006). Quantum game theory explanation of disjunction effects. In R. Sun, N. Miyake, Eds. Proceedings of the 28th Annual Conference of the Cognitive Science Society, pp. 131- 135. Mahwah, NJ: Erlbaum.

Busemeyer, J. R., Wang, Z., & Townsend, J. T. (2006). Quantum dynamics of human decision-making. Journal of Mathematical Psychology, 50, 220-241.

Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. (2011). A quantum theoretical explanation for probability judgment errors. Psychological Review, 118, 193-218.

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

Festinger, L. (1957). A Theory of Cognitive Dissonance. Stanford Univ. Press, Stanford.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. Trends in Cognitive Sciences, 14, 357-364.

Grover, L. K. (1997). Quantum mechanics helps in searching for a needle in a haystack. Physical Review Letters, 79, 325-328.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

Khrennikov, A. Y. (2010). Ubiquitous quantum structure: From psychology to finance. Springer.

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.

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

Moore, D. W. (2002). Measuring new types of question-order effects. Public Opinion Quarterly, 66, 80-91.

Myung, I. J. (2000). The importance of complexity in model selection. Journal of Mathematical Psychology, 44 (1) , 190-204.

Oaksford, M. & Chater, N. (2007). Bayesian rationality: The probabilistic approach to human reasoning. Oxford: Oxford University Press.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. Psychological Review, 109, 472-491.

Pothos, E. M. & Busemeyer, J. R. (2009). A quantum probability explanation for violations of 'rational' decision theory. Proceedings of the Royal Society B, 276, 2171-2178.

Pothos, E. M. & Busemeyer, J. R. (in press). Open peer commentary. The fallacy of normativism: falling in love with ourselves. A case for limited prescriptive normativism. Behavioral and Brain Sciences.

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

Shiffrin, R. S. and Busemeyer, J. R. (2011) Model selection applied to quantum probability models. In Proceedings of the 32nd Annual Conference of the Cognitive Science Society, LEA: Mahwah, NJ.

Tenenbaum, J. B, Kemp, C., Griffiths, T. L., & Goodman, N. (2011). How to grow a mind: statistics, structure, and abstraction. Science, 331, 1279-1285.

Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207–232.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124–1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjuctive fallacy in probability judgment. Psychological Review, 90, 293-315.

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

Yukalov, V., & Sornette, D. (2010). Decision theory with prospect interference and entanglement. Theory and Decision, 70, 283-328.