Journal of Mathematical Psychology 82 (2018) 73–83

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Are quantum-like Bayesian networks more powerful than classical Bayesian networks?☆

Catarina Moreira *, Andreas Wichert

Instituto Superior Técnico, INESC-ID, Av. Professor Cavaco Silva, 2744-016 Porto Salvo, Portugal

ARTICLE INFO

Article history:
Received 26 June 2017
Received in revised form 4 November 2017

Keywords:
Latent Variables
Quantum cognition
Bayesian networks
Quantum-like models

ABSTRACT

Recent works in the literature have proposed quantum-like Bayesian networks as an alternative decision model to make predictions in scenarios with high levels of uncertainty. Despite its promising capabilities, there is still some resistance in the literature concerning the advantages of these quantum-like models towards classical ones.

In this work, we developed a Classical Latent Bayesian network model and we compared it against its quantum counterpart: the quantum-like Bayesian network. The comparison was done using a well known Prisoner's Dilemma game experiment from Shafir and Tversky (1992), in which the classical axioms of probability theory are violated during a decision, and consequently the game cannot be simulated by pure classical models. In the end, we concluded that it is possible to simulate these violations using the Classical Latent Variable model, but with an exponential increase in its complexity. Moreover, this classical model cannot predict both *observed* and *unobserved* conditions from Shafir and Tversky (1992) experiments. The quantum-like model, on the other hand, is shown to be able to accommodate both situations for *observed* and *unobserved* events in a single model, making it more suitable and more general for these types of decision problems.

1. Introduction

The task of determining human judgments under uncertainty has got increasing attention in the scientific literature in the last decade (Moreira & Wichert, 2016b). More specifically, several models that are capable to predict or explain human decisions that are inconsistent with the laws of classical probability theory and logic (Crosson, 1999; Kuhberger, Komunska, & Josef, 2001; Lambdin & Burdsal, 2007; Tversky & Shafir, 1992) have been recently proposed. These models turn to quantum probability to explain human decision-making and are part of a new emerging discipline called Quantum Cognition (Busemeyer, 2015; Wang, Busemeyer, Atmanspacher, & Pothos, 2013).

Recent research shows that quantum-based probabilistic models are able to explain and predict scenarios that cannot be explained by pure classical models (Bruza, Wang, & Busemeyer, 2015; Busemeyer & Wang, 2015). However, there is still a big resistance in the scientific literature to accept these quantum-based models. Many researchers believe that one can model scenarios that violate the laws of probability and logic through classical probabilistic decision models that are often used in machine learning (Murphy, 2012). These violations of the laws of probability theory are hard to explain through classical theory and can have different types: violations to the Sure Thing Principle (Savage, 1954), disjunction/conjunction errors (Tversky & Kahneman, 1983), Ellsberg (Ellsberg, 1961)/Allais (Allais, 1953) paradoxes, order effects (Sudman & Bradburn, 1974), etc.

To accommodate these violations, several quantum-like models have been proposed in the literature. Note that, the term *quantum-like* is simply the designation that it is used to refer to any model that is applied in the domains outside of physics and that use the mathematical formalisms of quantum mechanics, abstracting them from any physical meaning and interpretations.

Although, the quantum cognition field is recent in the literature, there have been several different quantum-like models proposed in the literature. These models range from dynamical models (Busemeyer, Wang, & Lambert-Mogiliansky, 2009; Busemeyer, Wang, & Townsend, 2006; Pothos & Busemeyer, 2009), which make use of unitary operators to describe the time evolution since a participant is given a problem (or asked a question), until he/she makes a decision, to models that are based on contextual probabilities (Aerts & Aerts, 1994; Khrennikov, 2009b; Yukalov & Sornette, 2011). Quantum-like dynamical models have

☆ This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013 and through the PhD grant SFRH/BD/92391/2013. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

* Corresponding author.
E-mail addresses: catarina.p.moreira@tecnico.ulisboa.pt (C. Moreira), andreas.wichert@tecnico.ulisboa.pt (A. Wichert).

https://doi.org/10.1016/j.jmp.2017.11.003
0022-2496/

also been proposed in the literature to accommodate violations to the Prisoner's Dilemma Game (Pothos & Busemeyer, 2009), study the evolution of the interaction of economical agents in markets (Haven & Khrennikov, 2013; Khrennikov, 2009a) or even to specify a formal description of dynamics of epigenetic states of cells interacting with an environment (Asano et al., 2013). On the other hand, quantum-like models based on contextual probabilities, explore the application of complex probability amplitudes to define contexts that can interfere with the decision-maker (Khrennikov, 2005b, 2009b, 0000). For a survey about the applications of quantum-like models for the Sure Thing Principle, the reader can refer to Moreira and Wichert (2016b).

In the literature, it is clear and acceptable that simple and pure probabilistic models cannot accommodate human decisions that violate the laws of classical probability theory and logic (Busemeyer & Bruza, 2012). But can a more complex classical model simulate the paradoxical findings reported in the literature? In order to answer this question, we propose the application of latent variables in classical models to accommodate these paradoxical findings. By latent variables, we mean random variables that are *hidden*, that is, they cannot be directly measured in an experimental setting, but can be indirectly inferred from experimental data. These variables bring great advantages to cognitive models, because many observed variables can be condensed into a smaller number of hidden variables, enabling a dimensionality reduction of the model. For instance, in Psychology or Social Sciences, one can use latent variables to summarise the influence of several variables, such as beliefs, personality, social attitudes, etc., over the concept of *h* uman behaviour (Bollen, 2002; Griffiths, Steyvers, & Tenenbaum, 2007).

A well-known classical model that can include such dependencies is the Bayesian network (Pearl, 1988). This model represents relationships between random variables (such as causal and conditional dependencies) in an acyclic probabilistic graphical structure. Bayesian networks are powerful inference models that have been successfully applied over the years in different fields of the literature, mainly in artificial intelligence, genetics, medical decision-making, economics, etc.

In this work, we developed a classical Bayesian network that makes use of *latent variables* and we compared it against its quantum counterpart, the quantum-like Bayesian network, which was previously proposed in Moreira and Wichert (2016a). In the end, we conclude that it is possible to simulate the violations to the Sure Thing Principle using the classical Bayesian network with latent variables with an exponential increase in its complexity, however this model cannot predict both *observed* and *unobserved* experimental conditions from Shafir and Tversky (1992). On the other hand, the quantum-like model is shown to be able to accommodate both situations for *observed* and *unobserved* events in a single and general model. Note that the Sure Thing Principle is a concept widely used in game theory and was originally introduced by Savage (1954). This principle is fundamental in Bayesian probability theory and states that if one prefers action A over B under state of the world X, and if one also prefers A over B under the complementary state of the world X, then one should always prefer action A over B even when the state of the world is unspecified.

This manuscript is organised as follows. In Section 2, we introduce a general definition for latent variables. In Section 3, we present the prisoner's dilemma game and several works of the literature that report experiments, which violate the Sure Thing Principle in this game. In Section 4, we propose a classical Bayesian network model that makes use of Latent Variables to accommodate the paradoxical findings of the prisoner's dilemma game. In Section 5, it is introduced the quantum-like Bayesian network proposed in the work of Moreira and Wichert (2016a) as an alternative model to accommodate the several paradoxical findings reported in the literature. In Section 6, we make a discussion about the complexity involved in exact probabilistic inferences over classical and quantum-like Bayesian networks. We end this work with Section 7, which summarises the main points of this work: that the quantum-like Bayesian network model poses advantages towards the classical model with latent variables, since it can simulate both *observed* and *unobserved* phenomena in a single network, in contrast with the classical model requires extra hidden nodes (contributing to a decrease in efficiency) and cannot accommodate both *observed* and *unobserved* experimental conditions in a single model.

2. Latent variables

Most of the times, the data that is recorded (or observed) does not provide all the information that is needed to model a decision scenario. In these situations, latent variables are used to model complex patterns that we do not have the complete data for.

There is not a general and formal definition for latent variables (Bollen, 2002). Since it is a concept that is widely used across different multidisciplinary areas, it can be defined differently according to its application. However, a very simple and informal definition can be given as variables that are not directly observed from data, but can be inferred using the information of the variables that were recorded (Anandkumar, Hsu, Javanmard, & Kakade, 0000). Instead of specifying concrete relationships between variables, latent variables enable the abstraction of these relationships allowing a more general representation, which can be inferred from the observed variables.

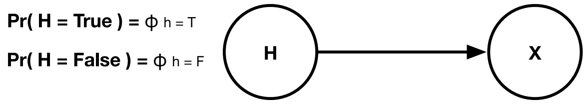
In this work, we will use latent variables in a probabilistic graphical model, more specifically in a Bayesian network. Generally speaking, a Bayesian network is an acyclic probabilistic graphical model, which provides an intuitive way of specifying probabilistic relationships and dependencies between random variables (Griffiths et al., 2007). These relationships are specified through a joint distribution over the set of all random variables in the model, and each node specifies conditional dependencies over its parent nodes. Under this representation, a random variable becomes latent when it is unobserved (or unknown), which suggests a *local independence* definition, according to Bollen (2002). When a latent variable is constant (for instance, a prior probability representing a person's cognitive bias towards some topic), the observed variables become independent. More formally, the independence between random variables and the latent variables is given by Eq. (1).

$$Pr(X_1, X_2, \ldots, X_n) = Pr(X_1|h)Pr(X_2|h) \cdots Pr(X_n|h). \quad (1)$$

Given a set of observed random variables $X_1, X_2, \ldots, X_n$ and some vector of latent (hidden) variables h, the joint probability $Pr(X_1, X_2, \ldots, X_n)$ corresponds to the product of the conditional probabilities of each random variable X_i over the associated latent variable, $Pr(X_1|h)Pr(X_2|h) \cdots Pr(X_n|h)$.

Consider Fig. 1. Suppose you have a parameterised acyclic probabilistic graphical model over the parameter ϕ. We will assume that node H represents a latent variable, because it is not directly observed (or it is hidden) for some given reason: it might be too expensive to collect its data, it might have been not recorded or we simply might not have access to the process generating the observed data.

Given a dataset of collected data D of size M, the above network consists in a tuple $\langle h[m], x[m] \rangle$, where h is parameterised instance of the latent variable H and x an instance of the random variable X. The likelihood (a measure similar to a probability, which provides



H	Pr(X = True | H)	Pr(X = False | H)
True	ϕ X = T | h = T	ϕ X = F | h = T
False	ϕ X = T | h = F	ϕ X = F | h = F

Fig. 1. Example of a Bayesian network with a latent variable H and a random variable X.

support for particular values of a parameter in a parametric model) of the network is given by the joint distribution:

$$\mathbb{L}(\phi : D) = \prod_{m=1}^{M} Pr(h[m], x[m] : \phi). \tag{2}$$

In a Bayesian network, the full joint probability distribution can be described in terms of the chain rule. So, Eq. (2) can be rewritten as:

$$\mathbb{L}(\phi : D) = \prod_{m=1}^{M} Pr(h[m] : \phi_H) \cdot Pr(x[m]|h[m] : \phi_{X|H}). \tag{3}$$

Note that Eq. (3) is composed by *two* terms, because the network in Fig. 1 has *two* random variables (more specifically, one random variable and one latent variable). For N random variables, this model would have N terms. Each term is called a *local likelihood function* and can determine how well a random variable can predict its parents (Murphy, 2012).

In Eq. (3), one can decompose the local likelihood function of the random variable X into two sets: one for each assignment that the random variable can take. In this case, for the sake of simplicity, it is assumed that X is a binary random variable that can be assigned the values *True* or *False*.

$$\begin{aligned} Pr(x[m]|h[m] : \phi_{X|H}) = & \prod_{m:h[m]=H_{True}} Pr(x[m]|h[m] : \phi_{X|H_{True}}) \\ & \cdot \prod_{m:h[m]=H_{False}} Pr(x[m]|h[m] : \phi_{X|H_{False}}). \end{aligned} \tag{4}$$

What is the probability $\prod_{m:h[m]=H_{True}} Pr(x[m]|h[m] : \phi_{X|H_{True}})$? Well, if we look at the conditional probability table in Fig. 1, we see that it is simply the product of the correspondent conditional probability distribution entry over all assignments of X. That means, when $X = true$, then $Pr(x[m]|h[m] : \phi_{X|H_{True}}) = \phi_{X=true|H=true}$. In the same way, when $X = false$, then $Pr(x[m]|h[m] : \phi_{X|H_{True}}) = \phi_{X=false|H=true}$. The same reasoning is applied for the second term of Eq. (4), but for H_{False}.

Given that we are making a product between all assignment of X, then $\prod_{m:h[m]=H_{True}} Pr(x[m]|h[m] : \phi_{X|H_{True}})$ can be rewritten as

$$\begin{aligned} & \prod_{m:h[m]=H_{True}} Pr(x[m]|h[m] : \phi_{X|H_{True}}) \\ & = \phi_{X=true|H=true}^{\#\langle h=true,x=true\rangle} \cdot \phi_{X=false|H=true}^{\#\langle h=true,x=false\rangle}, \end{aligned} \tag{5}$$

where the symbol # represents the cardinality. More specifically, $\#\langle h = true, x = false\rangle$ represents the number of instances where the latent variable H has the value *true* and the number of instances where the random variable X has the value *false*. From Eq. (5), it follows that:

$$\begin{aligned} \phi_{X=false|h=true} & = \frac{\#\langle h = true, x = false\rangle}{\#\langle h = true, x = true\rangle + \langle h = true, x = false\rangle} \\ & = \frac{\#\langle h = true, x = false\rangle}{\#\langle h = true\rangle}. \end{aligned} \tag{6}$$

Eq. (6) means that for a given network structure, we simply count the number of instances that each assignment X and H appear. The goal is to estimate the parameters of the latent variable in order to accommodate the violations of the several paradoxes reported over the literature. For the case of this work, to accommodate violations to the Sure Thing Principle.

In the next section, we present a decision scenario that has been widely reported over the literature as an example of violations to the Sure Thing Principle: the Prisoner's Dilemma Game (Shafir & Tversky, 1992).

3. The Prisoner's Dilemma game

The Prisoner's Dilemma game corresponds to an example of the violation of the Sure Thing Principle. In this game, there are two prisoners who are in separate solitary confinements with no means of speaking to or exchanging messages with each other. The police offer each prisoner a deal: they can either betray each other (defect) or remain silent (cooperate).

The dilemma of this game is the following. Taking into account the payoff matrix, the best choice for both players would be to cooperate. However, the action that yields a bigger individual reward is to defect. If player A has to make a choice, he has two options: if B has chosen to cooperate, the best option for A is to defect because he will be set free; if B has chosen to defect, then the best action for A is also to choose to defect because he will spend less time in jail than if he cooperates.

To test the veracity of the Sure Thing Principle under the Prisoner's Dilemma game, several experiments were performed in the literature in which three conditions were tested:

1. Participants were informed that the other participant chose to *defect*.
2. Participants were informed that the other participant chose to *cooperate*.
3. Participants had no information about the other participant's decision.

Table 1 summarises the results of several works in the literature that have performed this experiment using different payoffs. Note that although these results show that the majority of the players choose to *defect* in all three conditions of the experiment, there is still a significant percentage of players who choose to *cooperate*. Throughout the literature, it is considered that this significant percentage of players is violating the Sure Thing Principle (Pothos & Busemeyer, 2009) and, consequently, violating the law of total probability. In a classical setting, assuming that no prior information about the first player's preferences (neutral priors), it is expected that:

$$\begin{aligned} & Pr\ (\ P_2\ =\ Defect \mid P_1\ =\ Defect\ ) \geq Pr\ (\ P_2\ =\ Defect\ ) \\ & \geq Pr\ (\ P_2\ =\ Defect \mid P_1\ =\ Cooperate\ ). \end{aligned}$$

However, this is not consistent with the experimental results reported in Table 1. Note that $Pr(\ P_2 = Defect \mid P_1 = Defect\ )$ corresponds to the probability of the second player choosing the *Defect* action given that he knows that the first player chose to *Defect*. In Table 1, this corresponds to the entry Known to Defect. In the same manner, $Pr(\ P_2 = Defect \mid P_1 = Cooperate\ )$ corresponds to the entry Known to Cooperate. The observed probability during the experiments concerned with player 2 choosing to defect, $Pr\ (\ P2\ =\ Defect\ )$, corresponds to the unknown entry of Table 1

Table 1
Works of the literature reporting the probability of a player choosing to defect under several conditions.

Literature	Known to defect	Known to cooperate	Unknown	Classical probability
Shafir and Tversky (1992)	0.9700	0.8400	0.6300	0.9050
Li and Taplin (2002)[a]	0.8200	0.7700	0.7200	0.7950
Busemeyer, Matthew, and Wang (0000)	0.9100	0.8400	0.6600	0.8750
Hristova and Grinberg (0000)	0.9700	0.9300	0.8800	0.9500
Average	0.9175	0.8450	0.7225	0.8813

[a] Corresponds to the average of all seven experiments reported in the work of Li and Taplin (2002).

because there is no evidence regarding the first player's actions. Finally, the entry Classical Probability corresponds to the classical probability $Pr\ (\ P_2\ =\ Defect\ )$, which is computed through the law of total probability assuming neutral priors (a 50% chance of a player choosing either to cooperate or to defect):

$$Pr\ (P_2 = Defect) = Pr\ (P_1 = Defect)$$
$$\cdot\ Pr\ (P_2 = Defect|P_1 = Defect)$$
$$+\ Pr\ (P_1 = Cooperate) \cdot Pr\ (P_2 = Defect|P_1 = Cooperate)\ .$$

At this point, the reader might be thinking that the paradoxes reported in Table 1 can be easily modelled through classical probability models (Murphy, 2012) and for this reason there is no need to develop quantum probabilistic models for these problems. One can also think that the classical model fails to accommodate the paradox due to missing information. There can be some variables that might be influencing the participant's game and that cannot be measured directly (for instance, some arbitrary mental states, such as self-esteem, etc.). One can call these variables *hidden* and for that reason they can be modelled through latent variables. In the next section, we will present a classical Bayesian network model that uses latent variables to model the Prisoner's Dilemma game. This *hidden* state is introduced to show that by adding extra variables to the classical model, to capture the uncertainty of the player, is not enough to solve the paradox at hand.

4. Classical Bayesian network with latent variables

In order to discuss the idea that some hidden variable(s) might influence the participant's mental states leading to the paradoxical findings reported in Table 1, in this section we introduce a classical Bayesian network model with latent variables to model the Prisoner's Dilemma game.

A classical Bayesian network can be defined by a directed acyclic graph structure in which each node represents a different random variable from a specific domain and each edge represents a direct influence from the source node to the target node. The graph represents relationships between variables and each node is associated with a conditional probability table which specifies a distribution over the values of a node given each possible joint assignment of values of its parents. This idea of a node, depending directly from its parent nodes, is the core of Bayesian networks. Once the values of the parents are known, no information relating directly or indirectly to its parents or other ancestors can influence the beliefs about it (Koller & Friedman, 2009).

Consider Fig. 2, which represents a classical Bayesian network with a latent variable to model the Prisoner's Dilemma game.

In Fig. 2, $P1$ and $P2$ are both random variables. $P1$ represents the decision of the first player and $P2$ represents the second participant's decision (either defecting or cooperating). H is the *hidden state* or latent variable and represents some unmeasurable factor that can influence the players' decisions. For the sake of simplicity, let us assume that the latent variable has two states: *risk_seeking* and *risk_averse*. In the proposed classical model, the latent variable represents the personality of a player towards risk and suggests that a risk seeking player will tend to cooperate more, in contrast with a risk averse player will tend to defect more often. The main reason for making this assumption resides in two factors. First, over the literature, the prisoner's dilemma game is modelled under these two conditions when the purpose is to represent individual risk in decision-making tasks (Au, Lu, Leung, Yam, & Fung, 2011; Pothos & Busemeyer, 2009). Second, the complexity of the Bayesian network would grow exponentially and unnecessarily large with the incorporation of more random variables or with random variables with multiple assignments. Adding extra assignments to the latent variable will not bring any advantage in this decision-making problem as the reader will notice in the end of this section. Since a latent variable is a variable that is not directly measured, it can only be inferred from the observed data. The two random variables, $P1$ and $P2$, are the observed data from the experiment and are represented, respectively, by the functions $F(P1_i, H_j)$ and $G(P2_i, H_j)$ for $i \in \{defect, cooperate\}$ and $j \in \{risk\_averse, risk\_seeking\}$. These functions depend on the hidden and unmeasured variable H, which is parameterised over the parameter $K(H = j)$.

One can note that this model is in accordance with the definition of latent variables from Bollen (2002): if H is known, then the random variables P_1 and P_2 become independent:

$$Pr(P_1, P_2) = Pr(P_1|H).Pr(P_2|H).$$

The goal of this model is to find the parameter $K(H\ =\ j)$ from the observed experimental data such that all conditions of the Prisoner's Dilemma Game are satisfied. In other words:

1. When it is *known* that the first player chose to *defect*, then the participant should *defect*.
2. When it is *known* that the first player chose to *cooperate*, then the participant should *defect*.
3. When it is *not known* if the first player chose to *defect* or *cooperate*, then the second player should *cooperate*.

Assuming that m corresponds to the m_{th} element of the dataset D, the maximum likelihood estimate of this network with variable ϕ is given by the full joint probability distribution,

$$L\,(\phi : D) = \prod_{m=1} Pr\,(H\,[m]\,,P1\,[m]\,,P2\,[m] : \phi)\ . \tag{7}$$

Remember that H is the *hidden state* or latent variable that represents the personality of the player and is parameterised over the parameter $K(H\ =\ j)$, with $j\ \in\ \{risk\_averse, risk\_seeking\}$. $P1$ and $P2$ are both random variables. $P1$ represents the decision of the first player and $P2$ represents the second participant's decision (either defecting or cooperating). $P1$ and $P2$, are the observed data from the experiment and are represented, respectively, by the functions $F(P1_i, H_j)$ and $G(P2_i, H_j)$ for $i \in \{defect, cooperate\}$.

Given that we have three random variables in the network, the full joint probability distribution consists in the product of the nodes given their parents,

$$L\,(\phi : D) = \left(\prod_{m=1} Pr\,(H\,[m] : \phi_H)\right)$$

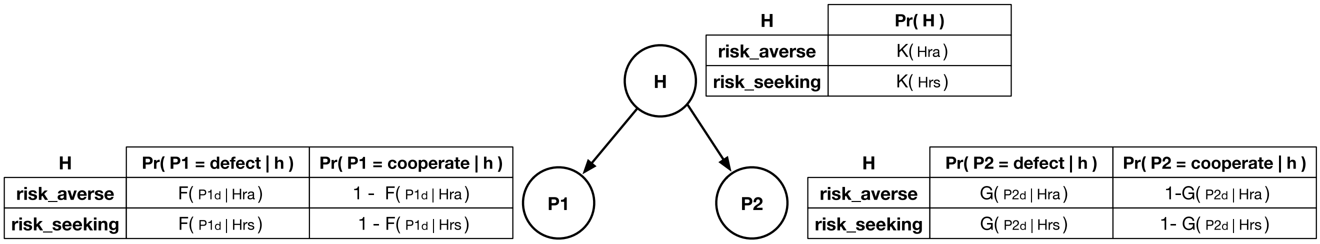


Fig. 2. A classical Bayesian network with a latent variables to model the Prisoner's Dilemma game. $P1$ and $P2$ are both random variables. $P1$ represents the decision of the first player and $P2$ represents the decision of the second player (either to cooperate or to defect). H is the *hidden state* or latent variable and represents some unmeasurable factor that can influence the participant's decisions.

$$\times \left(\prod_{m=1} Pr\left(P1\left[m\right]\left|H[m]\right. : \phi_{P1|H}\right)\right)$$
$$\times \left(\prod_{m=1} Pr\left(P2\left[m\right]\left|H[m]\right. : \phi_{P2|H}\right)\right). \tag{8}$$

The first term of Eq. (8) is the latent variable and corresponds to the prior probability about a player's personality. This hidden variable needs to be inferred from the observed data. The second and third terms correspond to player's P1 and P2 actions, respectively, and are the terms that we need to expand and compute. Since we are more interested in computing the probability of the second player, $P2$, we will do the calculations with this term. The calculations for $P1$ are performed in a similar way. Expanding the term about $P2$, the probability of the second player can be computed in terms of his personality (either risk seeking or risk averse), which is represented by the latent variable H.

$$\prod_{m=1} Pr\left(P2\left[m\right]\left|H[m]\right. : \phi_{P2|H}\right)$$
$$= \prod_{m:h[m]=risk\_averse} Pr(P2[m]|H[m] : \phi_{P2|H=risk\_averse}) \cdot \tag{9}$$
$$\prod_{m:h[m]=risk\_seeking} Pr(P2[m]|H[m] : \phi_{P2|H=risk\_seeking}).$$

Analysing each term, we see the latent variable can indeed be estimated by the random variables $P2$ and $P1$ through a function $G(P2 = i, H = j)$ and $F(P1 = i, H = j)$, for $i \in \{defect, cooperate\}$ and $j \in \{risk\_averse, risk\_seeking\}$, which corresponds to the number of times the assignments of H would appear together with each assignment of $P2$ and $P1$. More formally,

$$\prod_{H=risk\_averse} Pr\left(G2\left[m\right]\left|H\left[m\right]\right. : \phi_{G2|H=risk\_averse}\right)$$
$$= \left(\phi_{G2=defect|H=risk\_averse}\right) \cdot \left(\phi_{G2=cooperate|H=risk\_seeking}\right) \tag{10}$$

$$\phi_{P2=defect|H=risk\_averse} = \frac{\#\langle P2 = defect, H = risk\_averse\rangle}{\langle H = risk\_averse\rangle}$$
$$= G\left(P2 = defect, H = risk\_averse\right). \tag{11}$$

In the same way, we can make the same calculations for a risk seeking player:

$$\phi_{P2=defect|H=risk\_seeking} = \frac{\#\langle P2 = defect, H = risk\_seeking\rangle}{\langle H = risk\_seeking\rangle}$$
$$= G\left(P2 = defect, H = risk\_seeking\right). \tag{12}$$

For the random variable $P1$ the calculations are similar:

$$\phi_{P1=defect|H=risk\_averse} = \frac{\#\langle P1 = defect, H = risk\_averse\rangle}{\langle H = risk\_averse\rangle}$$
$$= F\left(P1 = defect, H = risk\_averse\right) \tag{13}$$

$$\phi_{P1=defect|H=risk\_seeking} = \frac{\#\langle P1 = defect, H = risk\_seeking\rangle}{\langle H = risk\_seeking\rangle}$$
$$= F\left(P1 = defect, H = risk\_seeking\right) \tag{14}$$

which leads to the parametrisation made in the Bayesian network model in Fig. 2. In the next section, we will try to find an estimation for the functions $F(P1 = i, H = j)$, $F(P1 = i, H = j)$ and for parameter $K(H = j)$ that can explain both *observed* and *unobserved* conditions for the Prisoner's Dilemma game.

4.1. Estimating the parameters

In order to find a classical Bayesian network that can explain the paradoxical results in Table 1, one needs to fit the conditional probabilities $F(P1 = i, H = j)$, $F(P1 = i, H = j)$ and the prior probability $K(H = j)$. In order to simulate the three conditions of the Prisoner's Dilemma game experiment, we need to satisfy two sets of conditions: one when the player does not know the decision of the first player and another one when the player knows the decisions of the first player.

These conditions can be taken from the computation of the full joint probability of the Bayesian network in Fig. 2. The full joint probability distribution of a Bayesian network corresponds to the multiplication of each assignment of a random variable with its parents. More specifically, for a set of random variables X that make up a Bayesian network, the full joint probability is given by Russel and Norvig (2010):

$$Pr(X_1, \ldots, X_n) = \prod_{i=1}^{n} Pr(X_i|Parents(X_i)). \tag{15}$$

We can specify the full joint probability of the Bayesian network in Fig. 2, through Eq. (15) by Table 2.

Note that throughout this work, we will address to *risk_averse* by *ra*, *risk_seeking* by *rs*, *defect* by *d* and *cooperate* by *c*.

The unobserved conditions correspond to the third condition of the experiment in which the participant does not know the decision of the first player. That is the same as computing the probability of a participant choosing to *defect* without further information: $Pr(P2 = defect)$. This results in the following conditions:

$$Pr(P2 = defect) = Pr(H = ra, P1 = d, P2 = d)$$
$$+ Pr(H = ra, P1 = c, P2 = d)+ \tag{16}$$
$$Pr(H = rs, P1 = d, P2 = d) + Pr(H = rs, P1 = c, P2 = d).$$

In the same way, one can specify the probability of the participant choosing to *cooperate*, $Pr(P2 = cooperate)$, as

Table 2
Full joint probability distribution for the general Bayesian network from Fig. 2, which models the Prisoner's Dilemma game. Note that *rs* stands for risk_seeking, *ra* for risk_averse, *d* for defect and *c* for cooperate.

H	$P1$	$P2$	$Pr(H, P1, P2)$
risk_averse	Defect	Defect	$K(H = ra)\, F(P1 = d, H = ra)\, G(P2 = d, H = ra)$
risk_averse	Defect	Cooperate	$K(H = ra)\, F(P1 = d, H = ra)\, G(P2 = c, H = ra)$
risk_averse	Cooperate	Defect	$K(H = ra)\, F(P1 = c, H = ra)\, G(P2 = d, H = ra)$
risk_averse	Cooperate	Cooperate	$K(H = ra)\, F(P1 = c, H = ra)\, G(P2 = c, H = ra)$
risk_seeking	Defect	Defect	$K(H = rs)\, F(P1 = d, H = rs)\, G(P2 = d, H = rs)$
risk_seeking	Defect	Cooperate	$K(H = rs)\, F(P1 = d, H = rs)\, G(P2 = c, H = rs)$
risk_seeking	Cooperate	Defect	$K(H = rs)\, F(P1 = c, H = rs)\, G(P2 = d, H = rs)$
risk_seeking	Cooperate	Cooperate	$K(H = rs)\, F(P1 = c, H = rs)\, G(P2 = c, H = rs)$

$$\alpha_1 = \frac{1}{\sum_{p \in \{d,c\}} \sum_{h \in \{ra,rs\}} K2(H2 = h)\, F2(OutP1 = d)\, G(P2 = p, OutP1 = d, H = h)}$$
$$\alpha_2 = \frac{1}{\sum_{p \in \{d,c\}} \sum_{h \in \{ra,rs\}} K2(H2 = h)\, F2(OutP1 = c)\, G(P2 = p, OutP1 = c, H = h)}.$$

Box I.

$$\begin{aligned} Pr(P2 = cooperate) &= Pr(H = ra, P1 = d, P2 = c) \\ &+ Pr(H = ra, P1 = c, P2 = c) + \\ &Pr(H = rs, P1 = d, P2 = c) + Pr(H = rs, P1 = c, P2 = c). \end{aligned} \tag{17}$$

Using the full joint probability in Table 2, one can rewrite Eqs. (16) and (17) and set the probabilities $Pr(P2 = defect)$, $Pr(P2 = cooperate)$ to the experimental values of the Prisoner's Dilemma game. For the case of the work of Shafir and Tversky (1992), then we should guarantee that $Pr(P2 = defect) = 0.63$, $Pr(P2 = cooperate) = 0.37$.

$$\begin{aligned} &Pr(P2 = defect) \\ &= K(H = ra)\, F(P1 = d, H = ra)\, G(P2 = d, H = ra) \\ &+ K(H = ra)\, F(P1 = c, H = ra)\, G(P2 = d, H = ra) \\ &+ K(H = rs)\, F(P1 = d, H = rs)\, G(P2 = d, H = rs) \\ &+ K(H = rs)\, F(P1 = c, H = rs) \\ &\times G(P2 = d, H = rs) = 0.63 \end{aligned} \tag{18}$$

$$\begin{aligned} &Pr(P2 = cooperate) \\ &= K(H = ra)\, F(P1 = d, H = ra)\, G(P2 = d, H = ra) \\ &+ K(H = ra)\, F(P1 = c, H = ra)\, G(P2 = d, H = ra) \\ &+ K(H = rs)\, F(P1 = d, H = rs)\, G(P2 = d, H = rs) \\ &+ K(H = rs)\, F(P1 = d, H = rs)\, G(P2 = d, H = rs) = 0.37. \end{aligned} \tag{19}$$

Eqs. (18) and (19) specify the *unobserved* conditions for the Prisoner's Dilemma Game and to satisfy them, we need to set their parameters in the following way (note that the following values represent one possible solution, however this solution is not unique).

$$\begin{aligned} &K(H = rs) = 0.5 \quad F(P1 = d, H = rs) = 0.1 \\ &G(P2 = d, H = rs) = 0.36 \\ &K(H = ra) = 0.5 \quad F(P1 = d, H = ra) = 0.9 \\ &G(P2 = d, H = ra) = 0.9. \end{aligned} \tag{20}$$

This means that there is indeed a classical model that explains the paradoxical findings of the Prisoner's Dilemma Game, which violate the laws of classical probability theory. However, we also need to satisfy the conditions when the player *knows* the decisions of the first player. For this, we will need to change the model. Consider the Bayesian network in Fig. 3, which represents a classical Bayesian network to model the observed conditions for the Prisoner's Dilemma Game. In this network, $OutP1$ and $P2$ are both random variables that represent the outcome (or decision) of the first player and the decision of the second player. $H2$ represents a latent unmeasurable variable that corresponds to the personality of the second player: either *risk averse* or *risk seeking*.

Since the second player will have access to the first player's decision, then we will need to make add a dependency between the nodes $OutP1$ and $P2$. Using the same line of thought, since we know the outcome of the first player, then we do not need any dependency between this node and the latent variable (since, for this experimental condition, it will only affect the second player). Using this model, the *observed* conditions for the Prisoners Dilemma game are given by the probabilities $Pr(P2|OutP1 = defect, H2)$, when the player is informed that the first player chose to *defect* and $Pr(P2|OutP1 = cooperate, H2)$, when the player is informed that the first player chose to *cooperate*.

$$\begin{cases} Pr(P2 = d|P1 = d) = \alpha_1[K2(H2 = ra)\, F2(OutP1 = d) \\ \quad \times G(P2 = d, OutP1 = d, H = ra) \\ \quad + K2(H2 = rs)\, F2(OutP1 = d) \\ \quad \times G(P2 = d, OutP1 = d, H = rs)] = 0.97 \\ Pr(P2 = c|P1 = d) = \alpha_1[K2(H2 = ra)\, F2(OutP1 = d) \\ \quad \times G(P2 = c, OutP1 = d, H = ra) \\ \quad + K2(H2 = rs)\, F2(OutP1 = d) \\ \quad \times G(P2 = c, OutP1 = d, H = rs)] = 0.03 \end{cases} \tag{21}$$

$$\begin{cases} Pr(P2 = d|P1 = c) = \alpha_2[K2(H2 = ra)\, F2(OutP1 = c) \\ \quad \times G(P2 = d, OutP1 = c, H = ra) \\ \quad + K2(H2 = rs)\, F2(OutP1 = c) \\ \quad \times G(P2 = d, OutP1 = c, H = rs)] = 0.84 \\ Pr(P2 = c|P1 = c) = \alpha_2[K2(H2 = ra)\, F2(OutP1 = c) \\ \quad \times G(P2 = c, OutP1 = c, H = ra) \\ \quad + K2(H2 = rs)\, F2(OutP1 = c) \\ \quad \times G(P2 = c, OutP1 = c, H = rs)] = 0.16. \end{cases} \tag{22}$$

In Eqs. (21) and (22), the variable α is the normalisation factor due to Naïve Bayes independence assumptions and is defined by α_1 and α_2 given in Box I.

In order to satisfy the *observed* conditions, we need to set their parameters in the following way (note that the following

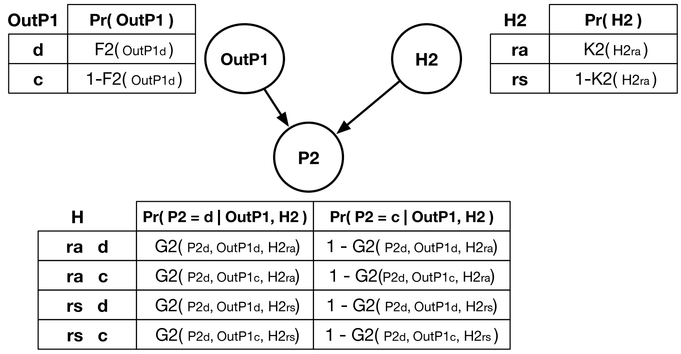


Fig. 3. Classical Bayesian network to model the observed conditions for the Prisoner's Dilemma Game. $OutP1$ and $P2$ are both random variables that represent the outcome (or decision) of the first player and the decision of the second player. The decisions can either be *defect*, which is represented by d or *cooperate*, represented by c. $H2$ represents a latent (hidden) unmeasurable variable that corresponds to the personality of the second player: either *risk averse* (ra) or *risk seeking* (rs).

values represent one possible solution, however this solution is not unique).

$$\begin{aligned} &K2(H = rs) = 0.5 \qquad F2(OutP1 = d) = 0.9 \\ &G2(P2 = d, OutP1 = d, H = rs) = 0.97 \\ &K2(H = ra) = 0.5 \qquad F2(OutP1 = c) = 0.1 \\ &G2(P2 = d, OutP1 = c, H = ra) = 0.84. \end{aligned} \tag{23}$$

This means that there is also a classical model that explains the observed findings of the Prisoner's Dilemma Game, which violate the laws of classical probability theory. But is there a single classical model that can accommodate both *observed* and *unobserved* conditions?

Let us go back to the Bayesian network in Fig. 2 (the one we used to compute the unobserved conditions). In order to satisfy the experimental results for the Prisoner's Dilemma game, one needs to find values for the parameters $K(H = j)$, $F(P1 = d, H = j)$ and $G(P2 = d, H = j)$, for $j \in \{ra, rs\}$, such that all *observed* and *unobserved* conditions are satisfied.

Making the calculations, one finds that the only way to satisfy both conditions, would be to set the parameters to:

$$\begin{aligned} &K(H = rs) = 0.9999 \qquad F(P1 = d, H = rs) = -3442.8 \\ &G(P2 = d, H = rs) = 0.6301 \\ &K(H = ra) = 0.0001 \qquad F(P1 = d, H = ra) = -3443.8 \\ &G(P2 = d, H = ra) = 0.3699. \end{aligned} \tag{24}$$

This is an impossible statement, because $F(P1 = d, H = ra) = -3443.8$ is violating the positivity axiom of classical probability theory. Moreover, the parameter $G(P2 = d, H = ra) = 0$ does not make sense, because it would imply that a player with a risk averse personality would always choose to *cooperate*, which is a contradiction. A risk averse player would prefer to *defect*, which is the action that leads to a higher utility. This shows that in a classical Bayesian network with latent variables, we can either find the conditional probability tables to accommodate the paradoxical findings of the Prisoner's Dilemma game or to accommodate the observed conditions. Satisfying both conditions in a single model is not possible.

4.2. Increasing the dimensionality of a classical Bayesian network

One can argue that adding another layer of hidden variables might solve the problem at hand and we would be able to simulate both *observed* and *unobserved* conditions. Although this line of thought is legitimate, it still does not solve the problem. Consider Fig. 4, which presents a classical Bayesian network. We have introduced a latent variable $H2$ that *joins* the models that can address the paradoxical findings and the models that can address the observed conditions of the Prisoner's Dilemma game. This means that by increasing the dimensionally of the model, we can obtain a network that takes into account both *observed* and *unobserved* conditions. In Fig. 4, $H1$ and $H2$ are latent variables that express both *unobserved* and *observed* conditions for the Prisoner's Dilemma game. Random variables $P1U$ and $P1$ represent the first player's decision according to the *unobserved* and *observed* conditions, respectively. Random variables $P2U$ and $P2$ represent the second player's decision according to the *unobserved* and *observed* conditions, respectively. The assignments ra stand for risk averse, rs risk seeking, d defect and c cooperate.

In order to address all conditions for the Prisoner's Dilemma game, we need to find the values for parameter ϕ_0 that would lead to the experimental outcomes of Shafir and Tversky (1992) work reported in Table 1. There are two possible ways to get a value for ϕ_0, but they both lead to a contradiction to the decision model.

1. Setting $\phi = 1$, would make $Pr(P2 = defect|P1 = defect) = 0.97$, $Pr(P2 = defect|P1 = cooperate) = 0.84$ and $Pr(P2U = defect) = 0.63$. This reflects the observations of Shafir and Tversky (1992) experiments, however we have two problems. First, setting a latent variable to a 100% probability goes against its definition, since we would be implying that this hidden variable that affects the player's decisions is always present. This leads to the second problem. Under such parameterisation, we would not be able to justify the probability $Pr(P2U = defect) = 0.63$, because we would always be under an *observed* condition (and P2U represents the actions of the second player under *unobserved* conditions).
2. Setting $\phi = 2/3$, which represents the experimental setup: two experiments for observed conditions, and one experiment for the unobserved condition. With this parameterisation, the only way to meet the results for the *observed* conditions, would be if we *know* that $H2 = obs$. That is, only by computing the probability $Pr(P2 = defect|P1 = defect, H2 = obs)$ we would obtain the experimental values from Shafir and Tversky (1992) work. This again is a contradiction, because a latent variable can never be used as a piece of absolute information during an inference process. Since it is a hidden variable, it is not measurable.

One should also take into account that by adding extra hidden variables to a Bayesian network model, one is exponentially increasing the complexity of the model. For N binary random variables, if no information is observed, we would need to compute a full joint probability distribution with 2^N entries. The Prisoner's Dilemma game is just a small decision scenario and to attempt to accommodate all experimental observations, we required 6 binary random variables, which leads to a full joint probability distribution with $2^6 = 64$ entries to be stored in memory. For more complex decision scenarios, these computations grow exponentially large and the inference process becomes intractable.

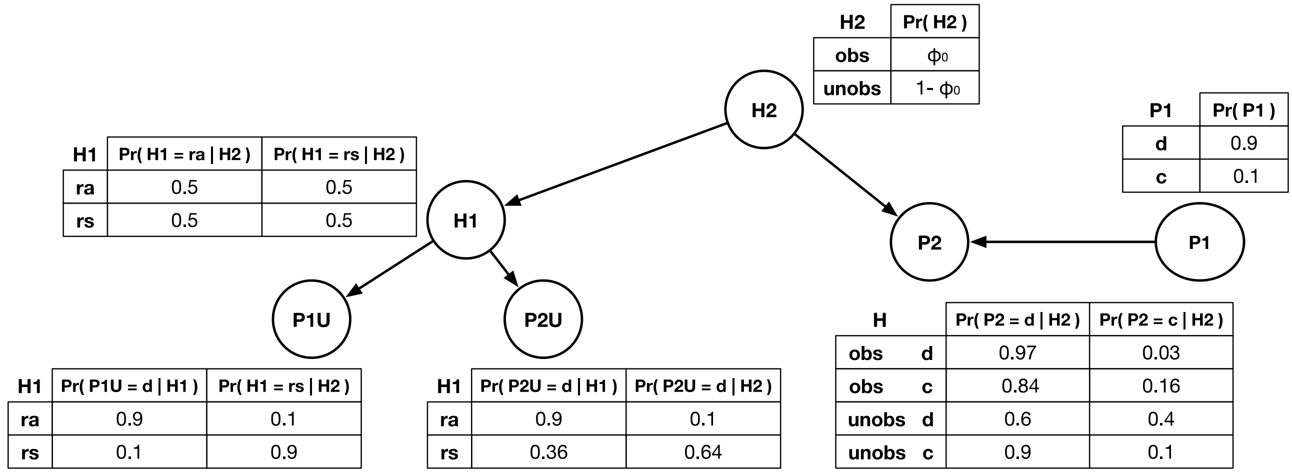


Fig. 4. A general classical Bayesian network with two latent variables, $H1$ and $H2$, to express both *unobserved* and *observed* conditions for the Prisoner's Dilemma game. Random variables $P1U$ and $P1$ represent the first player's decision according to the *unobserved* and *observed* conditions, respectively. Random variables $P2U$ and $P2$ represent the second player's decision according to the *unobserved* and *observed* conditions, respectively. The assignments *ra* stand for risk averse, *rs* risk seeking, *d* defect and *c* cooperate.

In the next section, we propose an alternative model based on quantum probability theory that can take into account both *observed* and *unobserved* conditions in general and compact way.

5. Quantum-like Bayesian networks as an alternative model

A more recent work from Moreira and Wichert (2014) suggested defining the quantum-like Bayesian network in the same way as a classical Bayesian network, but replacing real probability numbers by quantum probability amplitudes.

In this sense, we can build a quantum-like Bayesian network by applying Born's rule (Deutsch, 1988), that is by replacing the classical full joint probability distribution and the classical marginal probability distribution by quantum complex amplitudes. Then, we just apply the squared magnitude to the equation. The quantum-like full joint probability distribution is given by Eq. (25).

$$Pr(X_1, \ldots, X_n) = \left| \prod_{i=1}^{N} \psi_{(X_i | Parents(X_i))} \right|^2 . \tag{25}$$

Just like it is mentioned in the works of Moreira and Wichert (2014, 2016a), the general idea of a quantum-like Bayesian network is that, when performing probabilistic inference, the probability amplitudes of each assignment of the network are propagated and influence the probabilities of the remaining nodes, causing quantum interference effects to occur. In other words, *every* assignment of *every* node of the network is propagated until the node representing the query variable is reached.

By applying Born's rule, one can get the quantum counterpart of the classical marginal probability distribution. In other words, one can obtain a quantum-like version of the classical exact inference formula in the following way:

$$Pr(X|e) = \alpha \left| \sum_{y} \prod_{x=1}^{N} \psi_{(X_x | Parents(X_x), e, y)} \right|^2 . \tag{26}$$

Expanding Eq. (26), it will lead to the quantum interference formula:

$$Pr(X|e) = \alpha \left(\sum_{i=1}^{|Y|} \left| \prod_{x}^{N} \psi_{(X_x | Parents(X_x), e, y=i)} \right|^2 + 2 \cdot Interference \right)$$

$$Interference = \sum_{i=1}^{|Y|-1} \sum_{j=i+1}^{|Y|} \left| \prod_{x}^{N} \psi_{(X_x | Parents(X_x), e, y=i)} \right| \cdot \left| \prod_{x}^{N} \psi_{(X_x | Parents(X_x), e, y=j)} \right| \cdot \cos(\theta_i - \theta_j). \tag{27}$$

In the end, we need to normalise the final scores that are computed to achieve a probability value, because we do not have the constraints of double stochastic operators. In classical Bayesian inference, normalisation of the inference scores is also necessary due to assumptions made in Bayes rule. The normalisation factor corresponds to α in Eq. (27).

Note that, in Eq. (27), if one sets $(\theta_i - \theta_j)$ to $\pi/2$, then $\cos(\theta_i - \theta_j) = 0$, which means that the quantum Bayesian network collapses to its classical counterpart. That is, they can behave in a classical way if one sets the interference term to zero. Approaches to tune those parameters under a quantum-like Bayesian network approach are still an open research question, however some works have proposed to set these parameters though heuristic functions (Moreira & Wichert, 2015, 2016a, 2017). In the quantum-like Bayesian network, if there are many unobserved nodes in the network, then the levels of uncertainty are very high and the interference effects produce changes in the final likelihood of the outcomes, making it possible to explain the paradoxical results found in the literature.

The full joint probability distribution of the quantum-like Bayesian network in Fig. 5 is given by Table 3.

Using the quantum-like Bayesian network in Fig. 5, one can compute the probability of P2 = defect in the following way:

$$Pr(P2 = defect) = \alpha \left(\sum_{p \in P1} |\psi(P1 = p, P2 = d)|^2 + 2\, |\psi(P1 = d, P2 = d)| \times |\psi(P1 = c, P2 = d)|\, Cos\,(\theta_d - \theta_c) \right) \tag{28}$$

$$= \alpha\, (\psi(P1 = d, P2 = d) + \psi(P1 = c, P2 = d) + 2\, |\psi(P1 = d, P2 = d)| \times |\psi(P1 = c, P2 = d)|\, Cos\,(\theta_d - \theta_c)) . \tag{29}$$

The quantum interference term $\theta_d - \theta_c$ can either be set manually according to the experimental observations or it can be estimated

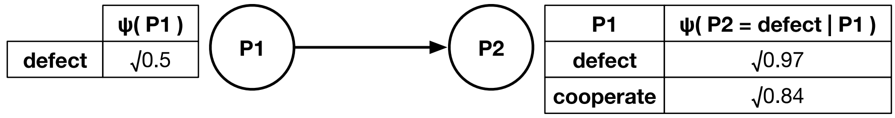


Fig. 5. Example of a quantum-like Bayesian network. The terms ψ correspond to quantum probability amplitudes. The variables $P1$ and $P2$ correspond to random variables representing the first and the second player, respectively.

Table 3
Full joint probability distribution table of the quantum-like Bayesian network in Fig. 5.

$P1$	$P2$	$\psi(P1, P2)$
Defect	Defect	$\psi(P1 = d)\psi(P2 = d \mid P1 = d) = \sqrt{0.5}\sqrt{0.97} = 0.6964$
Defect	Cooperate	$\psi(P1 = d)\psi(P2 = c\mid P1 = d) = \sqrt{0.5}\sqrt{0.03} = 0.1225$
Cooperate	Defect	$\psi(P1 = c)\psi(P2 = d\mid P1 = c) = \sqrt{0.5}\sqrt{0.84} = 0.6481$
Cooperate	Cooperate	$\psi(P1 = c)\psi(P2 = c\mid P1 = c) = \sqrt{0.5}\sqrt{0.16} = 0.2828$

using the similarity heuristic proposed in Moreira and Wichert (2016a). In this work, to be fair with the previously presented classical model, we will set the quantum parameter directly according to the experimental setting. That is, the quantum interference parameter is given by $\theta_d - \theta_c = 2.8151$. Continuing the calculations,

$$Pr(P2 = defect) = \alpha(\left(\sqrt{0.5}\sqrt{0.97}\right)^2 + \left(\sqrt{0.5}\sqrt{0.84}\right)^2 + 2\sqrt{0.5}\sqrt{0.97}\sqrt{0.5}\sqrt{0.84}\, Cos\,(2.8151)\,) = \alpha\; 0.05. \tag{30}$$

In the same way, we compute the probability of the second player choosing to cooperate:

$$\begin{aligned} &Pr(P2 = cooperate) \\ &\quad = \alpha \left(\sum_{p \in P1} |\psi(P1 = p, P2 = c)|^2 \right. \\ &\qquad + 2\,|\psi(P1 = d, P2 = c)| \\ &\qquad \left. \times\; |\psi(P1 = c, P2 = c)|\, Cos\,(\theta_d - \theta_c) \right) \end{aligned} \tag{31}$$

$$Pr(P2 = cooperate) = \alpha(\left(\sqrt{0.5}\sqrt{0.03}\right)^2 + \left(\sqrt{0.5}\sqrt{0.16}\right)^2 + 2\sqrt{0.5}\sqrt{0.03}\sqrt{0.5}\sqrt{0.16}\, Cos\,(2.8151)\,) = \alpha\; 0.0294. \tag{32}$$

Making the calculations we would end up with the probabilities

$$Pr(P2 = defect) = 0.63 \qquad Pr(P2 = cooperate) = 0.37,$$

which simulate the results obtained in the work of Shafir and Tversky (1992) described in Table 1 under unobserved events.

If it is known the action of Player 1, then the probability of Player 2 choosing to defect is given by:

$$Pr(P2 = defect|P1 = defect) = \alpha|\psi(P1 = d, P2 = d)|^2 = 0.97$$

$$Pr(P2 = defect|P1 = cooperate) = \alpha|\psi(P1 = c, P2 = d)|^2 = 0.84.$$

This shows that the quantum-like Bayesian network is a general a suitable model to be applied in scenarios with high levels of uncertainty that violate the laws of classical probability theory, since it can account for both *observed* and *unobserved* events without requiring hidden variables.

Table 4 shows which values the quantum interference parameter must have to simulate several works over the literature that report violations to the Sure Thing Principle. A method to estimate these parameters through heuristic functions has been proposed in the work of Moreira and Wichert (2016a).

In Table 4, the column $\theta_i - \theta_j$ represents the value of the quantum parameters of Eq. (27) that need to be set to explain the paradoxical findings reported in the several works of the literature. The column $Pr(P2 = Defect)$ corresponds to the probability of the second player choosing the action *defect* (the unobserved condition). The columns $Pr(P2 = Defect\ |P1 = Defect)$ and $Pr(P2 = Defect\ |P1 = Cooperate)$ correspond to the probability of the second player choosing to defect, given that it is known that the first player chose the actions *defect* and *cooperate*, respectively (the observed conditions).

It is important to note that the quantum-like Bayesian network is just an example of a quantum-like model that is able to express both *observed* and *unobserved* conditions of the prisoner's dilemma game. However, it is not the only model capable of achieving this. We chose this model, because it represents the quantum counterpart of the classical Bayesian network with latent variables that we are proposing in this work. There are other quantum-like models, which are able to accommodate the paradoxical situations found in the prisoner's dilemma game. The most representative ones correspond to the application of the quantum dynamical model (Busemeyer et al., 2006; Pothos & Busemeyer, 2009) and the quantum-like approach (Khrennikov, 2009b).

The quantum dynamical model takes into account time evolution to express the participants' beliefs and decisions throughout time using unitary operators (Busemeyer et al., 2009; Wang & Busemeyer, 2016). This dynamical representation also enables the simulation of dissonance effects. That is, the participants might have been confronted by some information that conflicted with his/her existing beliefs. In the end, the Quantum Dynamical model shows that quantum probability is a very general framework and can also accommodate both *observed* and *unobserved* experimental conditions of the prisoner's dilemma game.

The quantum-like approach (Khrennikov, 1999, 2001, 2003, 2005a, 0000) is another example of a quantum model that can represent both paradoxical findings and observed conditions in a single model. The quantum-like approach makes use of *contexts* to model decision scenarios. The context relates to the circumstances of setting an event in terms of which it can be fully understood, clarifying the meaning of the event. More specifically, it is a complex of conditions under which a measurement is performed. For instance, in domains outside of physics, such as cognitive science, one can have mental contexts. In social sciences, we can have a social context. And the same idea is applied to many other domains, such as economics, politics, game theory, biology, etc. These contexts will enable the representation of interferences between quantum states, which will allow the accommodation of the paradoxical findings.

Table 4
Analysis of the quantum θ_x parameters computed for each work of the literature in order to reproduce the *observed* and *unobserved* conditions of the Prisoner's Dilemma Game.

Literature	$\theta_i - \theta_j$	$Pr(P2 = \text{Defect})$	$Pr(P2 = \text{Defect} \mid P1 = \text{Defect})$	$Pr(P2 = \text{Defect} \mid P1 = \text{Cooperate})$
Shafir and Tversky (1992)	2.8151	0.6300	0.9700	0.8400
Li and Taplin (2002)[a]	3.3033	0.7200	0.8200	0.7700
Busemeyer et al. (0000)	2.9738	0.6600	0.9100	0.8400
Hristova and Grinberg (0000)	2.8255	0.8800	0.9700	0.9300

[a] Corresponds to the average of all seven experiments reported.

6. Discussion about the complexity of classical and quantum-like Bayesian networks

It is straightforward that quantum-like Bayesian networks suffer the same problem of the exponential increase of complexity (expressed as the dimension of the state space) as the classical Bayesian networks. Indeed, in what concerns the complexity of the inference problem, Bayesian networks (either classical or quantum-like) will always be *NP-Hard*. This means that exact inference on Bayesian networks are part of a class of problems that are extremely hard for a computer to solve, because it takes an exponential number of computational steps to perform the computations. The hardness of the exact inference comes precisely in the computation of the full joint probability distribution, which takes at most $2^N - 1$, computational steps assuming that all random variables of the network are binary, for N being the number of nodes in the network. This gives a complexity of $O(2^N)$. If random variables are not binary, then the exact inference process becomes even worse with a complexity of $O(M^N)$, where M is the number of assignments that the random variables can have.

This work shows that in order for a classical Bayesian network to reproduce the *unobserved* experimental conditions that lead to the violations of the Sure Thing Principle, one needs to add an extra random variable to the model. This means that just to accommodate these paradoxical findings, exact inference on a classical network would require $2^{(N+1)}$ computational steps, and with the disadvantage of only being able to explain the paradoxical results of Shafir and Tversky (1992) experiments. To explain the results of the *observed* conditions, we would need another network representation of the problem. To sum up, to explain Shafir and Tversky (1992) experiments, we would need two Bayesian networks (one for the observed conditions and another for the paradoxical findings), where one of the networks would need an increase of complexity to explain the violations of the Sure Thing Principle.

The quantum-like Bayesian network, on the other hand, has advantages towards the classical network for two reasons: (1) only one model is required to reproduce both *observed* and *unobserved* conditions of Shafir and Tversky (1992) experiments and (2) the quantum interference terms can be computed in quadratic time with an addition of $m(m+1)/2 - m$ operations, where m is the size of the marginal probability distribution using the heuristic proposed in the previous study of Moreira and Wichert (2016a). This means that, to accommodate violations to the Sure Thing Principle, the computation of the heuristic and the quantum interference terms is much less costly than performing inference on a classical Bayesian network with an additional random variable: $O(2^N + m(m+1)/2 - m) < O(2^{(N+1)})$. Although inferences are NP-hard in Bayesian networks, it is still possible (and manageable) to perform inferences in scenarios for a large N (for instance, a finite N repeated prisoner's dilemma game, medical decision-making, etc.). So far, preliminary research shows that quantum-like inferences over Bayesian networks with 21 nodes (which result in a full joint probability table with 6 291 456 entries) is manageable (Moreira, Haven, Sozzo, & Wichert, 0000).

Summarising, exact inference on a quantum-like Bayesian network can predict the *observed* and *unobserved* conditions of Shafir and Tversky (1992) experiments in a single model, without the need of adding extra random variables and with a reduced complexity when compared with the classical model.

7. Conclusion

The application of quantum principles to model decision-making scenarios emerged in the scientific literature as a way to explain and understand human behaviour in situations with high levels of uncertainty that lead to the violation of the classical laws of probability theory and logic. However, many researchers are still resistant in accepting the promising advantages of these quantum-like models towards modelling decision-making scenarios. Many times it is argued that classical models can simulate these decision scenarios under high levels of uncertainty adding extra variables to the model that are not directly observed through data. That is, by including extra latent (or hidden) variables it was believed that the model could represent uncertainty in the same way as in a quantum-like model, despite the complexity of the classical model.

In this work, we study this classical conception and make a mathematical comparison between a classical Bayesian network with Latent variables with the quantum-like Bayesian network previously proposed in the work of Moreira and Wichert (2016a). Latent Variables can be defined as variables that are not directly observed from data, but they can be inferred using the information of the variables that were recorded. For a complete dataset and given the full network structure, latent variables can be estimated by simply counting how many times they can be inferred from each assignment of the observed random variables.

We also validated these two models against the Prisoner's Dilemma Game, which is highly mentioned in the Cognitive Psychology domain Busemeyer et al. (0000), Conte et al. (2007), Crosson (1999), Hristova and Grinberg (0000), Li and Taplin (2002) and Shafir and Tversky (1992). This experiment is suitable to validate both classical and quantum-like models, because it violates the classical laws of classical probability theory and, consequently, it cannot be simulated by pure classical models.

Experimental results show that, although the classical model with latent variables could explain the paradoxical findings under the Prisoner's Dilemma game, the same model could not simulate the choice of the player when a piece of evidence was given, that is, when it was *known* which action the first player chose. This leads to the dilemma: either one creates a classical model just to account for observed evidence or one creates the model just to explain the paradoxical findings. Of course, one could argue that adding another extra Hidden latent variable to the network, one could re-estimate the conditional probability tables to account for both observed and unobserved phenomena. However, one must also take into account the exponential increase of complexity of the model. For N binary random variables, if no information is observed, we would need to compute a full joint probability distribution with 2^N entries, which means that the number of computations required grow exponentially large and the inference process becomes intractable.

On the other hand, it was already shown in previous literature that the quantum-like Model can account for both *observed* and *unobserved* phenomena in a single model with a low error percentage (Moreira & Wichert, 2016a). Since no extra nodes are incorporated in the network (when compared to the classical latent variable model), then the quantum model has also the advantage of a reduced complexity towards its classical counterpart.

Summarising, in this work we conclude that the quantum-like Bayesian network model poses advantages towards the classical model with latent variables, since it can simulate both *observed* and *unobserved* phenomena in a single network, in contrast with the classical model would need extra hidden nodes (contributing to a decrease in efficiency) and cannot simulate both phenomena on the same model.

References

Aerts, D., & Aerts, S. (1994). Applications of quantum statistics in psychological studies of decision processes. *Journal of Foundations of Science*, *1*, 85–97.

Allais, M. (1953). Le comportement de l'homme rationel devant le risque: Critique des postulats et axiomes de l'cole americaine. *Econometrica*, *21*, 503–546.

Anandkumar, A., Hsu, D., Javanmard, A., & Kakade, S. M. (0000). Learning linear Bayesian networks with latent variables. In *Proceedings of the 30th international conference on machine learning*.

Asano, M., Basieva, I., Khrennikov, A., Ohya, M., Tanaka, Y., & Yamato, I. (2013). A model of epigenetic evolution based on theory of open quantum systems. *Journal of Systems and Synthetic Biology*, *7*, 161–173.

Au, W. T., Lu, S., Leung, H., Yam, P., & Fung, J. M. Y. (2011). Risk and prisoner's dilemma: A reinterpretation of coombs' re-parameterization. *Journal of Behavioral Decision-Making*, *25*, 476–490.

Bollen, K. (2002). Latent variable in psychology and the social sciences. *Annual Review in Psychology*, *53*, 605–634.

Bruza, P., Wang, Z., & Busemeyer, J. (2015). Quantum cognition: a new theoretical approach to psychology. *Trends in Cognitive Sciences*, *19*, 383–393.

Busemeyer, J. (2015). Cognitive science contributions to decision science. *Cognition*, *135*, 43–46.

Busemeyer, J., & Bruza, P. (2012). *Quantum model of cognition and decision*. Cambridge University Press.

Busemeyer, J., Matthew, M., & Wang, Z. (0000). A quantum information processing explanation of disjunction effects, cognitive science. In *Proceedings of the 28th annual conference of the cognitive science society*.

Busemeyer, J., & Wang, Z. (2015). What is quantum cognition, and how is it applied to psychology? *Current Directions in Psychological Sciences*.

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

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

Conte, E., Todarello, O., Federici, A., Vitiello, F., Lopane, M., Khrennikov, A., et al. (2007). Some remarks on an experiment suggesting quantum like behavior of cognitive entities and formulation of an abstract quantum mechanical formalism to describe cognitive entity and its dynamics. *Chaos, Solitons & Fractals*, *31*, 1076–1088.

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

Deutsch, D. (1988). Quantum theory of probability and decisions. *Proceedings of the Royal Society of London, Series A (Mathematical and Physical Sciences)*, *455*, 3129–3137.

Ellsberg, D. (1961). Risk, ambiguity and the savage axioms. *Quaterly Economics*, *75*, 643–669.

Griffiths, T., Steyvers, M., & Tenenbaum, J. (2007). Topics in semantic representation. *Psychological Review*, *114*, 211–244.

Haven, E., & Khrennikov, A. (2013). *Quantum social science*. Cambridge University Press.

Hristova, E., & Grinberg, M. (0000). Disjunction effect in prisonner's dilemma: Evidences from an eye-tracking study. In *Proceedings of the 30th annual conference of the cognitive science society*.

Khrennikov, A. (1999). Classical and quantum mechanics on information spaces with applications to cognitive psychological, social and anomalous phenomena. *Foundations of Physics*, *29*, 1065–1098.

Khrennikov, A. (2001). Linear representations of probabilistic transformations induced by context transitions. *Journal of Physics A: Mathematical and General*, *34*, 9965–9981.

Khrennikov, A. (2003). Representation of the kolmogorov model having all distinguishing features of quantum probabilistic model. *Physics Letters A*, *316*, 279–296.

Khrennikov, A. (2005a). Linear and nonlinear analogues of the schrödinger equation in the contextual approach to quantum mechanics. *Journal of Doklady Mathematics*, *72*, 791–794.

Khrennikov, A. (2005b). The principle of supplementarity: A contextual probabilistic viewpoint to complementarity, the interference of probabilities, and the incompatibility of variables in quantum mechanics. *Journal of Foundations of Physics*, *35*, 1655–1693.

Khrennikov, A. (2009a). Classical and quantum-like randomness and the financial market. In *Coping with the complexity of economics*. Springer.

Khrennikov, A. (2009b). Quantum-like model of cognitive decision-making and information processing. *Journal of BioSystems*, *95*, 179–187.

Khrennikov, A. (0000). From classical statistical model to quantum model through ignorance of information. In *Proceedings of the third conference on the foundations of information science*.

Khrennikov, A. Quantum-like representation of macroscopic configurations. In *Proceedings of the 3rd international symposium on quantum interactions*.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. The MIT Press.

Kuhberger, A., Komunska, D., & Josef, P. (2001). The disjunction effect: Does it exist for two-step gambles? *Organizational Behavior and Human Decision Processes*, *85*, 250–264.

Lambdin, C., & Burdsal, C. (2007). The disjunction effect reexamined: Relevant methodological issues and the fallacy of unspecified percentage comparisons. *Organizational Behavior and Human Decision Processes*, *103*, 268–276.

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

Moreira, C., Haven, E., Sozzo, S., & Wichert, A. (0000). A quantum-like analysis of a real financial scenario: The Dutch bank loan application (extended abstract). In *Proceedings of the 13th econophysics colloquium*, (p. 35), http://www.ec2017.org/wp-content/uploads/2017/06/27-06-2017_boa.pdf.

Moreira, C., & Wichert, A. (2014). Interference effects in quantum belief networks. *Applied Soft Computing*, *25*, 64–85.

Moreira, C., & Wichert, A. (2015). The synchronicity principle under quantum probabilistic inferences. *NeuroQuantology*, *13*, 111–133.

Moreira, C., & Wichert, A. (2016a). Quantum-like Bayesian networks for modeling decision-making. *Frontiers in Psychology*, *7*.

Moreira, C., & Wichert, A. (2016b). Quantum probabilistic models revisited: the case of disjunction effects in cognition. *Frontiers in Physics: Interdisciplinary Physics*, *4*, 1–26.

Moreira, C., & Wichert, A. (2017). Exploring the relations between quantum-like bayesian networks and decision-making tasks with regard to face stimuli. *Journal of Mathematical Psychology*, *78*, 86–95.

Murphy, K. (2012). *Machine learning: A probabilistic perspective*. MIT.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publishers.

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

Russel, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach*. (3rd ed.). Pearson Education.

Savage, L. (1954). *The foundations of statistics*. John Wiley.

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

Sudman, S., & Bradburn, N. (1974). *Response effects in surveys: A review and synthesis*. Aldine Publishers Co.

Tversky, A., & Kahneman, D. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

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

Wang, Z., & Busemeyer, J. (2016). Interference effects of categorization on decision-making. *Cognition*, *150*, 133–149.

Wang, Z., Busemeyer, J., Atmanspacher, H., & Pothos, E. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science*, *5*, 689–710.

Yukalov, V., & Sornette, D. (2011). Decision theory with prospect interference and entanglement. *Theory and Decision*, *70*, 283–328.