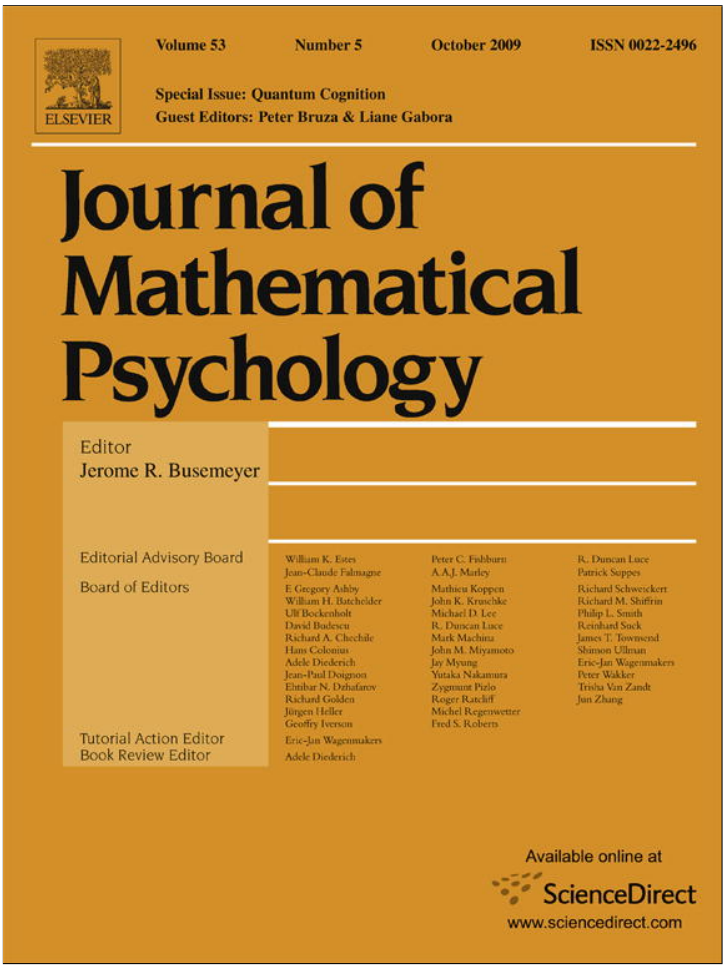
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Journal of Mathematical Psychology 53 (2009) 423–433

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Empirical comparison of Markov and quantum models of decision making[☆]

Jerome R. Busemeyer[a,*], Zheng Wang[b], Ariane Lambert-Mogiliansky[c]

[a] Indiana University, United States
[b] Ohio State University, United States
[c] PSE Paris School of Economics, France

ARTICLE INFO

Article history:
Received 20 June 2008
Received in revised form
23 February 2009
Available online 14 May 2009

Keywords:
Markov processes
Quantum probability
Categorization
Decision making
Interference effects
Double stochasticity

ABSTRACT

There are at least two general theories for building probabilistic–dynamical systems: one is Markov theory and another is quantum theory. These two mathematical frameworks share many fundamental ideas, but they also differ in some key properties. On the one hand, Markov theory obeys the law of total probability, but quantum theory does not; on the other hand, quantum theory obeys the doubly stochastic law, but Markov theory does not. Therefore, the decision about whether to use a Markov or a quantum system depends on which of these laws are empirically obeyed in an application. This article derives two general methods for testing these theories that are parameter free, and presents a new experimental test. The article concludes with a review of experimental findings from cognitive psychology that evaluate these two properties.

© 2009 Elsevier Inc. All rights reserved.

Markov theory is a general mathematical framework for describing probabilistic–dynamical systems, which is commonly used in all areas of cognitive science (Townsend & Ashby, 1983). For example, it is the mathematical framework that underlies random walk/diffusion models of decision making, or stochastic models of information processing, or multinomial processing tree models of memory retrieval, as well as many other applications. Quantum theory provides an alternative general mathematical framework for describing probabilistic–dynamical systems (Gudder, 1988). In fact, these two mathematical frameworks share many fundamental ideas, but they also differ in some key properties. The purpose of this article is to identify and empirically test some basic properties that distinguish these theories. It will be shown that, on the one hand, Markov theory obeys the law of total probability, but quantum theory does not; on the other hand, quantum theory obeys the doubly stochastic law, but Markov theory does not. Therefore, the decision about whether to use a Markov or a quantum system depends on which of these laws are empirically obeyed in an application. The research described below examines whether there are cognitive situations that violate the Markov properties and obey the quantum properties instead.

The article is organized as follows. First, we describe a new experiment on categorization and decision making that provides an empirical data set suitable for comparing the two systems. Second, Markov and quantum models are presented in a parallel fashion to examine their fundamental similarities and differences, and the basic testable properties of each model are presented. Finally, the empirical findings are used to test and evaluate the basic properties of each system.

1. Categorization — Decision making paradigm

Townsend (Townsend, Silva, Spencer-Smith, & Wenger, 2000) introduced a new paradigm to study the interactions between categorization and decision making, which we discovered is highly suitable for testing Markov and quantum models. We recently replicated and extended their earlier work using the following experimental methods. On each trial, participants were shown pictures of faces, which varied along two dimensions (face width and lip thickness). Two different distribution of faces were used: On average a 'narrow' face distribution had a narrow width and thick lips; on average a 'wide' face distribution had a wide width and thin lips. The participants were asked to categorize the faces as belonging to either a 'good' guy or 'bad' guy group, and/or they were asked to decide whether to take an 'attack' or 'withdraw' action. The primary manipulation was produced by using the following four test conditions, presented in different trials, to each participant. In the C-then-D condition, participants made a categorization followed by an action decision; in the D-then-C condition, participants made an action decision followed by a

☆ This material is based upon work supported by the National Science Foundation SES Grant No.s 0817965 and 0818277.

* Corresponding address: Indiana University, Department of Psychological and Brain Sciences, 1101 E. 10th street, 47405-7007 Bloomington, IN, United States.
E-mail address: jbusemey@indiana.edu (J.R. Busemeyer).

0022-2496/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2009.03.002

Author's personal copy

categorization; in the C-Alone condition, participants only made a categorization; and finally in the D-Alone condition, participants only made an action decision.

This paradigm provides a straightforward test of a classical information processing model called general recognition theory (Ashby & Townsend, 1986). According to this theory, on each trial, the presentation of a face produces a perceptual image, which is represented as a point within a two-dimensional (face width, lip thickness) perceptual space. Furthermore, each point in the perceptual space is assigned to a 'good' guy (denoted G) or 'bad' guy (denoted B) category response label; and at the same time, each point is also assigned a 'withdraw' (denoted W) or 'attack' (denoted A) action. Let $G\&W$ represent the set of points that are assigned to the 'good' guy category and the 'withdraw' action; and analogous definitions apply to form the sets $G\&A$, $B\&W$, and $B\&A$. Thus the probability of categorizing the face as a 'good' guy and taking a 'withdraw' action, denoted $\Pr(G\&W)$, equals the probability of sampling a face that belongs to the $G\&W$ set of points; the other three probabilities, $\Pr(G\&A)$, $\Pr(B\&W)$, and $\Pr(B\&A)$, are determined in an analogous manner. The marginal probability of taking a 'defensive' action is determined by the *law of total probability*:

$$\Pr(A) = \Pr(G\&A) + \Pr(B\&A) = \Pr(G)\Pr(A|G) + \Pr(B)\Pr(A|B). \quad (1)$$

The categorization–decision paradigm provides a simple test of the law of total probability. In particular, this paradigm allows one to compare (i) the probability of taking an 'attack' action obtained from the D-Alone condition with the total probability computed from the C-then-D condition, and (ii) the probability of making a 'good' guy categorization obtained from the C-Alone condition with the total probability computed from the D-then-C condition. Townsend et al. (2000) reported chi-square tests of (i) at the .05 significance level. They found that with narrow faces, 38 out of 138 participants produced statistically significant deviations; with wide faces, 34 out of 138 did so. These numbers are much higher than what is expected by chance alone (using a significance level at .05 the expected number is only $(.05)(138) = 6.9$).

How can we explain these violations of the law of total probability? Quantum probability theory may provide an answer because it does not necessarily obey this law. However, it obeys a different law, called double stochasticity. Therefore, we also need to examine the latter property within the categorization–decision paradigm. However, Townsend et al. (2000) only reported the chi-square statistics, and not the original choice proportions, and so it is not possible to evaluate the double stochasticity using their report. For this reason, we conducted a new experiment that replicated and extended their original findings, which is described below.

1.1. Method

Participants

In total, 26 undergraduate students from a Midwest University participated in the study. Of them, 84.6% were female; 92.31% were Caucasian; and 7.69% were African American. The average age was 21.23 (SE = .22). The participants were awarded with extra course credit. In addition, they were motivated to perform well in order to win bonus credit points on top of the base participation credit.

Stimuli

Each participant viewed a series of faces on a computer. Thirty-four headshots of Caucasian men were selected from the photograph book Heads (Kayser, 1985). The photos were digitally scanned and then altered using the Adobe Photoshop CS3 software to enhance the manipulated features: (1) half of the faces were narrow and with thick lips, and the other half were wide with thin lips; (2) the ratio between the face width (between temples) and the face height (between the chin and the top of the head) was .83 (SD = .05) for the wide faces and .60 (SD = .02) for the narrow faces; and (3) the ratio between the lip thickness and the lip width was .22 (Std = .05) for thin lips and .38 (SD = .05) for thick lips. Other features of the faces were controlled: all faces had neither hair nor makeup; all showed calm and non-emotional facial expression; and all were 15 cm from chin to the top of the head when presented on the experiment computers. Using a separate group of students ($N = 6$), a pretest asking about the features of the face showed that the manipulation of the face shapes and lip thickness was significant.

As noted above, the faces vary according to two salient cues, the face shape (wide vs. narrow) and the lip thickness (thin vs. thick), which were probabilistically related to the categories. As in the study by Townsend et al. (2000), the wide faces with thin lips had a 60% chance to be assigned to the 'Adok' group, and 40% to be a 'Lork'. The narrow faces had a 40% chance to be the 'Adok' group, and 60% to be a 'Lork'. Then, for the Adoks, they had a 70% chance for the correct action to be 'friendly' or 'withdraw' and 30% to be 'hostile' or 'attack'. For the Lorks, they had a 30% chance for the correct action to be 'withdraw' and 70% to be 'attack'. Participants were given full information about the cues and the associated probabilities (not numerically, but in natural language) during the instruction at the beginning of an experimental session. (See the Appendix for exact instructions).

Experimental procedure

Participants were run in groups of two to seven, with office cube separation to minimize distraction. Each participant completed two experimental sessions during the same daytime hours on two consecutive days. Each experiment session lasted around 40 min and was implemented in MediaLab. The first session included Blocks 1–4 and the second session included Blocks 5–7.

For Blocks 1–6, each block included 34 trials with one trial for each face stimulus. The 34 trials were randomly divided into two groups, with one group including eight wide faces and nine narrow faces, and the other group including the rest of the 17 faces. Then, these two groups were randomly assigned to the two conditions: categorization then decision (C-then-D) and decision then categorization (D-then-C). The 17 trials within a group were randomized. Which group was presented first in a block was also randomized.

During each C-then-D trial, first, a face was presented at the center of the computer screen for 10 s while the categorization question was asked at the top of the screen: "Is this face an Adok or a Lork?" After the participant clicked a response key, the face remained on the screen but the categorization question was replaced immediately by the action decision question: "Would you be friendly or defensive?" After a response was made, a feedback page was presented on the screen for three seconds. This page included feedback for both responses. For an "Adok" categorization response, if the face was pre-assigned as Adok, the feedback would be "Yes! It was an Adok". If the face was pre-assigned as Lork, it would be "No! It was not an Adok, but a Lork". For a "Lork" response, the feedback follows the same logic and format. For a "Friendly" action decision response, if the Adok was pre-assigned to be 'friendly' or 'withdraw', the feedback would be "Yes! You are friendly to a friendly Adok. The Adok handed you $ 20". If it was pre-assigned to be 'hostile' or 'attack', the feedback would say: "No! You were friendly to a hostile Adok. You were mugged". By the same token, feedback was given to other response combinations. To facilitate the feedback information processing, the key words of "yes" and "no" were highlighted using green and red color responsively, a small picture of the face (6 cm from the chin to the top of the head) was presented with the feedback, and a similar sized picture showing the action consequences (i.e., 20 dollars, being mugged). For both questions, the participant had

Author's personal copy

up to 10 s to make a response using the "A" or "L" keys marked on the keyboard. If the participant failed to click either of these two keys within 10 s, a window popped up saying "The time limit for this question has passed". A missing data was recorded by the MediaLab program. The D-then-C trials were the same as the C-then-D trials except for the order of the two questions.

Trials in Block 7 were different from those in the other blocks. It included 68 trials, with two trials for each face—one asking the categorization question and the other asking the action decision question. Again, each question was given up to 10 s to answer, but the feedback page duration was reduced to two seconds since the information was simpler. All trials were randomized for each participant.

For all blocks, after the feedback was presented at the end of a trial, the computer asked: "Are you ready for the next trial?" To proceed, the participants needed to click a "continue" key marked on the keyboard. This allowed the participants to pace themselves through trials to reduce possible fatigue effects.

1.2. Results

Recall that participants were given instructions about the statistical relations between face types and categories, and between categories and actions. As a result, there were no systematic changes in choices across the first six blocks. Therefore, the choice proportions for each person were pooled across the first six blocks. The proportions for the C-then-D and D-then-C conditions are based on 51 trials per person times 26 persons, resulting in a sample size of $N = 1326$. The proportions for the D-Alone and C-Alone conditions are based on 17 trials per person times 26 persons, resulting in a sample size of $N = 442$.

The main results of the experiment are summarized in Table 1.1. The first group of three rows compare results from the C-then-D versus D-Alone conditions; the next group of three rows compare the results from the D-then-C versus C-Alone condition. The third group of three rows re-analyze the data from the D-then-C condition assuming a C-then-D order (which is discussed in more detail later in this section). The first column of the table indicates that the results were computed separately for trials using each type of face (wide versus narrow). The next four columns (2 through 5) indicate the probabilities of category and action responses, and the column labeled *TP* contains the total probability computed from the previous four columns. The last column indicates the choice probability from the last block of the experiment when only a single response was made. The difference between the *TP* column and the last column indicates the deviation from the prediction of the law of total probability.

First compare the *TP* versus $\Pr(A)$ columns obtained from the C-then-D versus D-Alone condition in the first group of three rows (labeled G1) in Table 1.1. The law of total probability is practically satisfied for the wide faces (.37 versus .39), but it is violated for the narrow faces (.59 versus .69). In fact, the proportion of 'attack' actions in the D-Alone condition (.69) is higher than both proportions (conditioned on categorization) obtained in the C-then-D condition for narrow faces (.41, .63). We computed the difference between *TP* and $\Pr(A)$ for each participant for narrow faces and found that 77% (20 out of 26) participants produced a difference in this direction. Furthermore, the t-test based on the difference scores for narrow faces was statistically significant ($t(25) = 2.54$, $SE = .034$, $p = .018$, two tailed).

Next consider the results obtained from the D-then-C versus C-Alone condition in the second group of three rows (labeled G2) in Table 1.1. Once again the law of total probability is generally satisfied. However, these results may be misleading for the following reason. Townsend et al. (2000) originally pointed out that the C-then-D processing order is more natural than the D-then-C order. In other words, participants may implicitly categorize faces first under the D-then-C condition, but report them second according to instructions. If this is true, then we would not expect any violations of the law of total probability. This is because we would be estimating the same probability of making a categorization *first* for both the D-then-C condition and the C-Alone condition. Townsend et al. (2000) checked the assumption that participants implicitly categorized first in the D-then-C condition by re-analyzing this joint frequency data assuming a C-then-D order. They found that this re-analysis very closely reproduced their findings from the C-then-D order. They concluded that under the D-then-C instructions, participants mentally processed the information in the more natural C-then-D order and simply reported these mental computed results in the D-then-C order.

Table 1.1
Experimental results.

G1	C-then-D					D-Alone
Type	$\Pr(G)$	$\Pr(A\|G)$	$\Pr(B)$	$\Pr(A\|B)$	*TP*	$\Pr(A)$
W	.84	.35	.16	.52	.37	.39
N	.17	.41	.83	.63	.59	.69
G2	D-then-C					C-Alone
Type	$\Pr(A)$	$\Pr(G\|A)$	$\Pr(W)$	$\Pr(G\|W)$	*TP*	$\Pr(G)$
W	.40	.73	.60	.84	.80	.79
N	.60	.15	.40	.28	.20	.23
G3	D-then-C re-analyzed as C-then-D					D-Alone
Type	$\Pr(G)$	$\Pr(A\|G)$	$\Pr(B)$	$\Pr(A\|B)$	*TP*	$\Pr(A)$
W	.80	.37	.20	.53	.40	.39
N	.20	.45	.80	.64	.60	.69
G4	Average C-then-D					D-Alone
Type	$\Pr(G)$	$\Pr(A\|G)$	$\Pr(B)$	$\Pr(A\|B)$	*TP*	$\Pr(A)$
W	.82	.36	.18	.53	.39	.39
N	.19	.43	.81	.63	.59	.69
QM	C-then-D					D-Alone
Type	$\Pr(G)$	$\Pr(A\|G)$	$\Pr(B)$	$\Pr(A\|B)$	*TP*	$\Pr(A)$
W	.82	.36	.18	.53	.39	.39
N	.19	.40	.81	.61	.57	.74

We followed Townsend et al. (2000) and re-analyzed the D-then-C frequencies to compute the probabilities shown in the third group of three rows (labeled G3) in Table 1.1. This re-analysis very closely replicates all of the results shown in the first group of three rows obtained from the C-then-D condition. This suggests that perhaps participants did mentally process the information in the C-then-D order (even though asked to do the D-then-C order), and simply reported these results in the D-then-C order. Note that the law of total probability is again practically satisfied for the wide faces (.40 versus .39), but it is once again violated for the narrow faces (.60 versus .69). As before, the proportion of 'attack' actions in the D-Alone condition (.69) is higher than both proportions (conditioned on categorization) obtained from the D-then-C condition for narrow faces (.45, .64). We again computed the difference between *TP* and $\Pr(D)$ for each participant for narrow faces, and found that 62% (16 out of 26) participants produced a difference in this direction. Furthermore, the t-test based on the difference scores for narrow faces was statistically significant ($t(25) = 2.19$, $SE = .037$, $p = .037$, two tailed). The correlation between the differences obtained from the C-then-D versus D-then-C conditions was $r = .92$ for narrow faces, which indicates a great deal of commonality between the two sets of results.

The fourth group of three rows (labeled G4) show the results when we average across the C-then-D and D-then-C conditions (as if the latter were processed in the C-then-D order). Again, we tested the differences $[TP - P(D)]$, now averaged across orders, for the narrow faces, and the t-test is again significant ($t(25) = 2.4$, $SE = .035$, $p = .02$, two tailed).

Author's personal copy

2. Comparison of Markov and quantum models

Below we empirically evaluate Markov and quantum models for the categorization–decision task using the average data for the C-then-D processing order (shown in the fourth group of rows in Table 1.1). First we present two-dimensional state models, and we show that both of these fail to explain the data for different reasons. Later we develop and compare more complex four-dimensional Markov and quantum models using the same data.

2.1. Two-dimensional Markov model

To fix ideas, we start by formulating a simple two-dimensional Markov model based on serial information processing and classic probabilities. According to this simple model, there is a set of two category states $C = \{|G\rangle, |B\rangle\}$ and a set of two action states $D = \{|A\rangle, |W\rangle\}$. If the C-then-D order is used to perform the task, then we assume that the decision maker starts in a state in C and makes a category response, and then transfers to one of the states in D to choose an action. The 2×1 column vector $P_I = \begin{bmatrix} p_G \\ p_B \end{bmatrix}$ represents the initial probabilities for the two category states (which depend on the type of face). This is a probability distribution, and so the entries in P_I are non-negative and sum to one. The 2×2 matrix $T = \begin{bmatrix} T_{AG} & T_{AB} \\ T_{WG} & T_{WB} \end{bmatrix}$ represents the probabilities of transiting from each category (column) state to each action (row) state (observed from the C-then-D condition). In particular, T_{ij} represents probability of transiting to state i from state j. This is a transition matrix and so the entries within each column are non-negative and the rows within each column sum to one. The final probability of taking each action (observed under the D-Alone condition) is given by the matrix product:

$$P_F = T \cdot P_I = \begin{bmatrix} P_{FA} \\ P_{FW} \end{bmatrix} = \begin{bmatrix} p_G \cdot T_{AG} + p_B \cdot T_{AB} \\ p_G \cdot T_{WG} + p_B \cdot T_{WB} \end{bmatrix}, \tag{2}$$

which is the same as the law of total probability given by Eq. (1).

This Markov model has three parameters $\{p_G, T_{AG}, T_{AB}\}$ that are estimated from four data points $\{\Pr(G), \Pr(A|G), \Pr(A|B), \Pr(A)\}$ for each type of face. For the wide faces, if we set $p_G = .82$, $T_{AG} = .36$, $T_{AB} = .53$, then we exactly recover $\Pr(G), \Pr(A|G), \Pr(A|B)$ and we predict $P_{FA} = .39$, which exactly matches the observed value $\Pr(A) = .39$. But for the narrow faces, if we set $p_G = .19$, $T_{AG} = .43$, $T_{AB} = .63$, then we exactly recover the data for $\Pr(G), \Pr(A|G), \Pr(A|B)$, but the model predicts $P_{FA} = .59$, whereas the observed value is significantly larger, $\Pr(A) = .69$. In summary, this simple Markov model fails because it requires the law of total probability, which is violated for the narrow faces.

2.2. Two-dimensional quantum model

Now we develop a simple two-dimensional quantum model based on serial information processing and probability amplitudes. According to quantum theory, amplitudes are basic, and probabilities are derived from their squared magnitudes (Feynman & Hibbs, 1965). In general, an amplitude is a complex number with a magnitude less than one: $q = r \cdot [\cos(\theta) + i \cdot \sin(\theta)]$, $i = \sqrt{-1}$, $r \in [0, 1]$, $\theta \in [0, 2\pi)$. Its squared magnitude equals $q \cdot q^* = (r \cdot [\cos(\theta) + i \cdot \sin(\theta)])(r \cdot [\cos(\theta) - i \cdot \sin(\theta)]) = r^2 \leq 1$. The parameter, θ, is called the phase, and if $\theta = 0$ or $\theta = \pi$, then q is real. (See the Appendix for a Hilbert space representation of this model).

Once again we assume that there are two category states $C = \{|G\rangle, |B\rangle\}$ and two action states $D = \{|A\rangle, |W\rangle\}$. If the C-then-D order is used to perform the task, then it is assumed that the decision maker starts in one of the states in C to make a category response, and transfers to one of the states in D to select an action. The 2×1 column vector $Q_I = \begin{bmatrix} q_G \\ q_B \end{bmatrix}$ represents the initial amplitudes for the category states (which depend on the type of face). The probability of initially observing the 'good' category equals $\Pr(G) = |q_G|^2$, and similarly $\Pr(B) = |q_B|^2$, and so the squared length of Q_I must equal one. The 2×2 matrix $U = \begin{bmatrix} U_{AG} & U_{AB} \\ U_{WG} & U_{WB} \end{bmatrix}$ represents the amplitudes for transiting from each category state to each action state. In quantum theory, this is a unitary matrix, $U^\dagger U = UU^\dagger = I$, so that it preserves inner products, which is required for a Hilbert space representation (see Appendix). The transition probabilities observed under the C-then-D condition are determined by the squared magnitudes of the entries in the unitary matrix. Specifically, $T_{ij}(U_{ij}) = \left|U_{ij}\right|^2$ is the probability of observing an action i given that a previous category response j was observed. A unitary matrix produces a transition matrix $T(U)$ that is doubly stochastic: both the rows and columns of the transition matrix generated by U sum to one (Peres, 1998). The final amplitude of taking each action under the D-Alone condition is given by the matrix product:

$$Q_F = U \cdot Q_I = \begin{bmatrix} Q_{FA} \\ Q_{FW} \end{bmatrix} = \begin{bmatrix} q_G \cdot U_{AG} + q_B \cdot U_{AB} \\ q_G \cdot U_{WG} + q_B \cdot U_{WB} \end{bmatrix}, \tag{3}$$

which can be interpreted as the *law of total amplitude* (Gudder, 1988). The probability of making an 'attack' decision in the D-Alone condition equals

$$\begin{aligned} |Q_{FA}|^2 &= (q_G \cdot U_{AG} + q_B \cdot U_{AB})(q_G \cdot U_{AG} + q_B \cdot U_{AB})^* \\ &= |q_G|^2 |U_{AG}|^2 + |q_B|^2 |U_{AB}|^2 + 2 \cdot |q_G||U_{AG}||q_B||U_{AB}| \cdot \cos(\theta), \end{aligned} \tag{4}$$

where θ is the phase of the complex number $(q_G \cdot U_{AG}) \cdot (q_B \cdot U_{AB})^*$. The law of total amplitude produces a probability (Eq. (4)) that violates the law of total probability (Eq. (1)) because of the interference term $2 \cdot |q_G| \cdot |U_{AG}| \cdot |q_B| \cdot |U_{AB}| \cdot \cos(\theta)$. If the interference term is zero (i.e., $\cos(\theta) = 0$), then the probability produced by the law of total amplitude (Eq. (4)) agrees with the law of total probability (Eq. (1)). However, if $\cos(\theta) > 0$, then Eq. (4) produces a higher probability than Eq. (1), as found in the data. In general, the interference can be positive, negative, or zero, depending on the phase θ.

This quantum model requires only three parameters $\{|q_G|^2, |U_{AG}|^2, \cos(\theta)\}$ to fit the four data points $\{\Pr(G), \Pr(A|G), \Pr(A|B), P(A)\}$. To see why, first note that double stochasticity implies that $|U_{WG}|^2 = 1 - |U_{AG}|^2 = |U_{AB}|^2$ and $|U_{WB}|^2 = 1 - |U_{WG}|^2 = |U_{AG}|^2$, and so the transition matrix produced by U has only one free transition probability $|U_{AG}|^2$. To fit the narrow face data, we set $|U_{AG}|^2 = .40$, which implies that $|U_{AB}|^2 = .60$ (which approximates the observed transition probabilities $\Pr(A|G) = .43$ and $\Pr(A|B) = .63$, respectively). We also set $|q_G|^2 = .19$, which exactly reproduces $\Pr(G) = .19$. Finally, if we set $\cos(\theta) = .333$, then Eq. (4) produces $|Q_{FA}|^2 = .69$, which exactly reproduces $\Pr(D)$ obtained from condition D-Alone. Note that the use of $\cos(\theta) = .333$ implies that complex amplitudes are required to fit the data.

Unfortunately, the quantum model fails to explain the wide face data. The reason is that the transition matrix observed under the C-then-D condition strongly violates double stochasticity, which is generated by U. Note that for the wide face data, the observed values, $\Pr(A|B) = .53$ and $1 - \Pr(A|G) = 1 - .36 = .64$ are far apart, but according to the quantum model, they are required to be equal. In summary, this two-dimensional quantum model fails because it requires the law of double stochasticity, which is violated for the wide face data. See Khrennikov (2007) for a two-dimensional amplitude theory that does not use unitary operators. Below we develop the four-dimensional models.

Author's personal copy

2.3. Four-dimensional Markov model

Next we develop a more sophisticated Markov model that assumes four combination states: $S = \{|G, A\rangle, |G, W\rangle, |B, A\rangle, |B, W\rangle\}$. For example, the state $|G, W\rangle$ represents a state in which the decision maker believes a face belongs to the 'good' category and the decision maker intends to take the 'withdraw' action. According to this model, the decision maker can switch from one state to another (or remain in the same state) at each moment in time. The 4×1 column vector P_I represents the initial probability distribution across the four states. Each row of P_I gives the probability of starting at one of the states in S, and the consecutive rows are symbolized as $(p_{GA}(0), p_{GW}(0), p_{BA}(0), p_{BW}(0))$. This initial distribution depends on the experimental conditions as follows.

Suppose the C-then-D order is used to perform the task. Following a 'good' guy categorization, the second two rows of the initial probability distribution are set to zero so that $(p_{GA} + p_{GW}) = 1$, and this initial probability distribution is denoted $P_I = P_G$. Following a 'bad' categorization, the first two rows of the initial probability distribution are set to zero so that $(p_{BA} + p_{BW}) = 1$, and this initial probability distribution is denoted $P_I = P_B$. For the D-Alone condition, it is assumed that there is some probability p_G of making an implicit 'good' categorization, and some probability $p_B = 1 - p_G$ of making an implicit 'bad' categorization so that the initial probability distribution is a mixture $P_I = (p_G \cdot P_G + p_B \cdot P_B)$.

After making a categorization, the decision maker deliberates for some period of time, t, to decide which action to take. This deliberation process is represented by a 4×4 transition matrix, $T(t)$, which is used to determine the new distribution across states after deliberation as follows: $P_F = T(t) \cdot P_I$. For example, $T_{ij}(t)$ determines the probability of transiting to state $i \in S$ from state $j \in S$ after deliberating for a time t. The transition probabilities in $T(t)$ are all non-negative real numbers and the rows within each column sum to one. It is well known (Bhattacharya & Waymire, 1990) that the transition matrix for a Markov process satisfies the Chapman–Kolmogorov equation $T(t + s) = T(t)T(s)$, and this implies that the transition matrix satisfies a differential equation called the Kolmogorov equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} T(t) = K \cdot T(t). \tag{5}$$

The matrix K in Eq. (5) is called the intensity matrix, which has non-negative off-diagonal entries and the rows within each column sum to zero in order to generate a transition matrix. The solution to Eq. (5) is the matrix exponential function $T(t) = e^{K \cdot t}$, which allows one to construct the transition matrix for any time point from the fixed intensity matrix. These intensities can be defined in terms of the evidence and payoffs for actions in the task. Later we will give an example of an intensity matrix for this task. However, we do not need to estimate a specific intensity matrix, because the theoretical property that we test holds for any transition matrix.

The 4×1 vector $P_F = T(t) \cdot P_I$ represents the probability distribution across states after deliberation, and the consecutive rows of P_F are symbolized as $(p_{GA}(t), p_{GW}(t), p_{BA}(t), p_{BW}(t))$. The probability of taking the 'attack' action equals $p_{GA}(t) + p_{BA}(t)$ and the probability of taking the 'withdraw' action equals $p_{GW}(t) + p_{BW}(t)$. It will be convenient to use an indicator matrix to select these choice probabilities. Define the 4×4 matrix M_A as a matrix with zeros everywhere except that a one is placed in the diagonal entry for row 1 and row 3. The vector produced by the matrix product, $M_A \cdot P_F$, contains the two probabilities of choosing the 'attack' action. We want to sum the elements of this vector, which can be done using 1×4 row vector $L = [1, 1, 1, 1]$. Then the probability of taking an 'attack' action equals $L \cdot (M_A \cdot P_F)$. Using these definitions we obtain the following probabilities for the C-then-D conditions:

$$\begin{aligned} \widehat{\Pr}(A|G) &= L \cdot (M_A \cdot T(t) \cdot P_G) \\ \widehat{\Pr}(A|B) &= L \cdot (M_A \cdot T(t) \cdot P_B). \end{aligned} \tag{6}$$

The probability for the D-Alone condition is equal to

$$\begin{aligned} \widehat{\Pr}(A) &= L \cdot M_A \cdot T(t) \cdot (p_G \cdot P_G + p_B \cdot P_B) \\ &= p_G \cdot L \cdot (M_A \cdot T(t) \cdot P_G) + p_B \cdot L \cdot (M_A \cdot T(t) \cdot P_B) \\ &= p_G \cdot \widehat{\Pr}(A|G) + p_B \cdot \widehat{\Pr}(A|B). \end{aligned} \tag{7}$$

Once again, we see that this more general Markov model satisfies the law of total probability, and consequently, it still fails to account for the violations found with the narrow faces. One might question the assumption that $p_G = \Pr(G)$ and $p_B = \Pr(B)$ in Eq. (7), where $\Pr(G)$ and $\Pr(B)$ are the observed probabilities of categorization from condition C-then-D, whereas p_G and p_B are implicit probabilities that operate under the D-Alone condition. However, even if we allow for this change in the probabilities of categorization across conditions, this Markov model fails because NO convex combination (weighted average) of $\Pr(A|G)$ and $\Pr(A|B)$ from the C-then-D condition can equal $\Pr(A)$ observed in condition D-Alone. This is because the latter exceeds both of the former, whereas a convex combination must lie in between.

2.4. Four-dimensional quantum model

Now we turn to a four-dimensional quantum model, and again we assume four combination states: $S = \{|G, A\rangle, |G, W\rangle, |B, A\rangle, |B, W\rangle\}$. As before, the state $|G, W\rangle$ represents a state in which the decision maker believes a face belongs to the 'good' guy category and the decision maker intends to take the 'withdraw' action. According to this model, the decision maker is in a superposition over states that evolves across time. The 4×1 column vector Q_I represents the initial amplitude distribution (quantum wave) across the four states. Each row of Q_I gives the amplitude for one of the states in S, and the consecutive rows are symbolized as $(q_{GA}(0), q_{GW}(0), q_{BA}(0), q_{BW}(0))$. Each amplitude is a complex number and the sum of squared magnitudes equals one ($|Q_I|^2 = 1$). As described for the Markov model, this initial distribution depends on the experimental conditions as follows.

Suppose the C-then-D order is used to perform the task. Following a 'good' categorization, the second two rows of the initial amplitude distribution are set to zero so that $|q_{GA}|^2 + |q_{GW}|^2 = 1$, and this initial amplitude distribution is denoted $Q_I = Q_G$. Following a 'bad' categorization, the first two rows of the initial distribution are set to zero so that $|q_{BA}|^2 + |q_{BW}|^2 = 1$, and this initial amplitude distribution is denoted $Q_I = Q_B$. For the D-Alone condition, it is assumed that the decision maker is superposed between these two categorization states so that the initial amplitude distribution is $Q_I = (q_G \cdot Q_G + q_B \cdot Q_B)$, with $|q_G|^2 + |q_B|^2 = 1$.

After making a categorization, the decision maker deliberates for some period of time, t, to decide which action to take. A 4×4 unitary matrix, $U(t)$, is used to determine the new superposition across states after deliberation as follows: $Q_F = U(t) \cdot Q_I$. For example, $U_{ij}(t)$ determines the amplitude for transiting to state $i \in S$ from state $j \in S$ after deliberating for a time t. The matrix $U(t)$ must satisfy $U(t)^\dagger U(t) = I$ to preserve inner products. The squared magnitude, $T_{ij} = |U_{ij}(t)|^2$, equals the probability of transiting to state $i \in S$ from state $j \in S$, and this transition matrix still satisfies double stochasticity. It is well known (Gudder, 1988) that the transition matrix for a quantum process also satisfies the amplitude version of the Chapman–Kolmogorov equation $U(t + s) = U(t)U(s)$, and this implies that the unitary matrix satisfies a

Author's personal copy

differential equation called the Schrödinger equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}U(t) = -i \cdot H \cdot U(t). \tag{8}$$

The matrix H in Eq. (8) is called the Hamiltonian matrix, which must be Hermitian, $H^{\dagger} = H$, in order to generate a unitary matrix. The solution to Eq. (8) is the matrix exponential function $U(t) = e^{-i \cdot H \cdot t}$, which allows one to construct the unitary matrix for any time point from the fixed Hamiltonian matrix. The elements of the Hamiltonian can be defined in terms of the evidence and payoffs for actions in the task. In a later section, we select a specific Hamiltonian to fit the observed choice probabilities. For now, we examine the general theoretical properties that hold for any unitary matrix.

The 4×1 vector $Q_F = U(t) \cdot Q_I$ represents the amplitude distribution across states after deliberation, and the consecutive rows of Q_F are symbolized as $(q_{GA}(t), q_{GW}(t), q_{BA}(t), q_{BW}(t))$. The probability of observing an 'attack' action equals $|q_{GA}(t)|^2 + |q_{BA}(t)|^2$ and the probability of observing the 'withdraw' action equals $|q_{GW}(t)|^2 + |q_{BW}(t)|^2$. Once again, it will be convenient to use an indicator matrix to select these choice probabilities. As before, define the 4×4 matrix M_A as a matrix with zeros everywhere except that a one is placed in the diagonal entry for row 1 and row 3. The vector produced by the matrix product, $M_A \cdot Q_F$, contains the two amplitudes for choosing the 'attack' action. The probability of the 'attack' action equals the squared length of this vector: $|M_A \cdot Q_F|^2$. Using these definitions we obtain the following probabilities for the C-then-D conditions:

$$\begin{aligned} Q_{A|G} &= M_A \cdot U(t) \cdot Q_G, \qquad \widehat{\Pr}(A|G) = |Q_{A|G}|^2 \\ Q_{A|B} &= M_A \cdot U(t) \cdot Q_B, \qquad \widehat{\Pr}(A|B) = |Q_{A|B}|^2. \end{aligned} \tag{9}$$

The probability of an 'attack' for the D-Alone condition is equal to

$$\begin{aligned} \widehat{\Pr}(A) &= |M_A \cdot U(t) \cdot (q_G \cdot Q_G + q_B \cdot Q_B)|^2 \\ &= |q_G \cdot (M_A \cdot U(t) \cdot Q_G) + q_B \cdot (M_A \cdot U(t) \cdot P_B)|^2 \\ &= |q_G \cdot Q_{A|G} + q_B \cdot Q_{A|B}|^2 \\ &= |q_G \cdot Q_{A|G}|^2 + |q_B \cdot Q_{A|B}|^2 \\ &\quad + 2 \cdot |q_G \cdot q_B \cdot (Q_{A|G}^{\dagger} \cdot Q_{A|B})| \cdot \cos(\theta) \\ &= p_G \cdot \widehat{\Pr}(A|G) + p_B \cdot \widehat{\Pr}(A|B) \\ &\quad + 2 \cdot |q_G \cdot q_B \cdot (Q_{A|G}^{\dagger} \cdot Q_{A|B})| \cdot \cos(\theta), \end{aligned} \tag{10}$$

where θ is the phase of the complex number $q_G \cdot q_B \cdot (Q_{A|G}^{\dagger} \cdot Q_{A|B})$. Once again, we see that the law of total probability fails for the quantum model whenever the interference term $2 \cdot |q_G \cdot q_B \cdot (Q_{A|G}^{\dagger} \cdot Q_{A|B})| \cdot \cos(\theta) \neq 0$. But what about the issue of double stochasticity? Although this model does satisfy the law of double stochasticity, this property is now defined in terms of the 4×4 transition matrix between combinations in S. Note that the event $\Pr(A|B)$ is no longer a transition between states in S for this model, and instead, it is the probability of a coarse measurement (projection on two states). This four-dimensional state space model no longer implies that $\Pr(A|B) = 1 - \Pr(A|G)$. In fact, in the next section, we show that the four-dimensional quantum model provides an adequate fit to all of the data.

2.5. Fitting models to data

In this section we describe more specifically how we apply the four-dimensional models to the categorization–decision experimental results. The models used here are based on an earlier application of these same models to a prisoner dilemma task in which players made choices based on either knowledge or no knowledge of an opponent's action (Pothos & Busemeyer, 2009). The capability of using the same models for these very different tasks demonstrates the generality of this modeling approach. Note that the specific Markov model presented below must fail because it must satisfy the law of total probability for any parameters, but we continue to present a specific version of it in order to identify the similarities and differences between this model and the quantum model.

Consider again the C-then-D order of processing. When a 'good' guy categorization is made, then initial state for the Markov model is denoted P_G with elements set to $p_{BA} = p_{BW} = 0$ and $p_{GA} = p_{GW} = .5$; similarly the initial state for the quantum model is denoted Q_G with elements set to $q_{BA} = q_{BW} = 0$ and $q_{GA} = q_{GW} = \sqrt{.5}$. When a 'bad' guy categorization is made, then the initial state for the Markov model is P_B with elements $p_{GA} = p_{GW} = 0$ and $p_{BA} = p_{BW} = .5$; similarly the initial state for the quantum model is Q_B with elements $q_{GA} = q_{GW} = 0$ and $q_{BA} = q_{BW} = \sqrt{.5}$. For the D-Alone condition, we assume that $P_I = p_G \cdot P_G + p_B \cdot P_B$, where $p_G = \Pr(G)$ is the probability of categorizing a face as a 'good' guy and $p_B = \Pr(B)$ is the probability of categorizing a face as a 'bad' guy; similarly, for the quantum model, we assume that $Q_I = \sqrt{p_G} \cdot Q_G + \sqrt{p_B} \cdot Q_B$, where p_G and p_B have the same values as defined for the Markov model. This assumption implies that the initial choice probabilities for action at time $t = 0$ are uniformly random when the decision maker has no time to deliberate about the best action to take after categorizing the face.

After categorizing the face, the decision maker deliberates for some period of time t and then chooses an action. For the Markov model, this is based on the final probability distribution $P_F = e^{t \cdot K} \cdot P_I$, where K is the intensity matrix; for the quantum model, this is based on the final amplitude distribution $Q_F = e^{-i \cdot t \cdot H} \cdot Q_I$, where H is the Hamiltonian matrix. (The matrix exponential function used in these calculations is available in most matrix algebra programming languages such as Matlab or R or Gauss.) These matrices (K for Markov and H for quantum) are assumed to be determined by two factors: one is used to change action preferences and the other is used to change category beliefs. More specifically, $K = K_d + K_b$ and $H = H_d + H_b$, where K_d and H_d change the action preferences, and K_b and H_b change the category beliefs, for the Markov and quantum models, respectively. Let us first consider the part that determines the preferences for actions (K_d and H_d).

For the Markov model, we set $K_d = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes K_{dG} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \otimes K_{dB}$ with $K_{dG} = \begin{bmatrix} -\mu_G & 1 \\ \mu_G & -1 \end{bmatrix}$ and $K_{dB} = \begin{bmatrix} -1 & \mu_B \\ 1 & -\mu_B \end{bmatrix}$. The submatrix K_{dG} transforms the initial probabilities toward favoring the 'withdraw' action when the decision maker believes the face is a 'good' guy, and the other submatrix K_{dB} transforms the probabilities toward favoring an 'attack' action when the decision maker believes the face is a 'bad' guy. The parameters, μ_G and μ_B, are assumed to be a function of the gains and losses produced by correctly or incorrectly choosing an action for each category. To help understand this part of the model, assume that a 'bad' guy category response is observed so that $P_I = P_B$. Then the probability of taking an 'attack' action equals

$$\begin{aligned} \widehat{\Pr}(A|B) &= L \cdot M_A \cdot e^{t \cdot K_d} \cdot P_B \\ &= \left(\frac{\mu_B}{1+\mu_B}\right) \cdot \left(1 - e^{-(\mu_B+1) \cdot t}\right) + \frac{1}{2} \cdot e^{-(\mu_B+1) \cdot t}. \end{aligned} \tag{11}$$

For $\mu_B > 0$, this probability starts at $\frac{1}{2}$ at $t = 0$ and grows monotonically toward an equilibrium $\left(\frac{\mu_B}{1+\mu_B}\right)$ as $t \to \infty$.

For the quantum model set $H_d = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes H_{dG} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \otimes H_{dB}$, where $H_{dG} = \begin{bmatrix} -\mu_G & 1 \\ 1 & \mu_G \end{bmatrix}$ and $H_{dB} = \begin{bmatrix} \mu_B & 1 \\ 1 & -\mu_B \end{bmatrix}$.

Author's personal copy

The submatrix H_{dG} rotates the initial amplitudes toward favoring the 'withdraw' action when the decision maker believes the face is a 'good' guy, and the other submatrix H_{dB} rotates the amplitudes toward favoring an 'attack' action when the decision maker believes the face is a 'bad' guy. To help understand this part of the model again, assume that a 'bad' guy response is observed so that $Q_I = Q_B$. Then the probability of taking an 'attack' action equals

$$\widehat{\Pr}(A|B) = |M_A \cdot e^{-i\cdot t\cdot H_d} \cdot Q_B|^2 = \frac{1}{2} + \left(\frac{\mu_B}{1+\mu_B^2}\right) \cdot \sin(t)^2. \tag{12}$$

For $-1 < \mu_B < 1$, this probability increases monotonically across time from $\frac{1}{2}$ at $t = 0$ to $\left(\frac{1}{2} + \frac{\mu_B}{1+\mu_B^2}\right)$ at $t = (\pi/2)$, and subsequently it oscillates between these minimum and maximum values. Empirically, choice probabilities in laboratory-based, decision making tasks increase monotonically across time (at least for relatively short decision times), and so a reasonable approach for fitting the model is to assume that a decision is reached within the interval $(0 < t < \pi/2)$ for the quantum model ($t = \pi/2$ would correspond to around 2 s for such tasks). Hereafter we will set $t = \pi/2$ for all calculations from the quantum model.

Using only K_d (for the Markov model) and H_d (for the quantum model) produces reasonable choice models for the C-then-D processing order when the category response is known. However, for condition D-Alone, when the category response is unknown, both models predict that the probability of defection is the average of the two known cases, which *fails* to explain the violations of the law of total probability. Although this is expected from the Markov model, this may be somewhat unexpected for the quantum model. The problem arises with the quantum model because in this special case $Q^{\dagger}_{A|G} \cdot Q_{A|B} = 0$ and there is no interference.

For the quantum model to explain violations of total probability, we need to introduce the idea of cognitive dissonance (Festinger, 1957): People tend to change their beliefs to be consistent with their actions. In the case of the category–decision paradigm, category beliefs tend toward consistency with the intended action. In other words, if the decision maker intends to 'attack' then he/she also tends to think that the face is a 'bad' guy. Thus both beliefs and action tendencies evolve and influence each other across time during the deliberation period. Next we show how these changes in beliefs can be included in both models. However, we also show that including the dissonance effect only helps the quantum model and does not change the basic predictions of the Markov model.

For the Markov model this dissonance effect can be produced by using

$$K_b = K_{bA} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + K_{bW} \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

where

$$K_{bA} = \begin{bmatrix} -\gamma & 1 \\ \gamma & -1 \end{bmatrix} \quad \text{and} \quad K_{bW} = \begin{bmatrix} -1 & \gamma \\ 1 & -\gamma \end{bmatrix}.$$

The submatrix K_{bA} changes beliefs toward a 'bad' guy category when the decision maker tends toward an 'attack' action, and the other submatrix K_{bW} changes beliefs toward a 'good' guy category when the decision maker tends toward a 'withdraw' action. For example, if $P_I' = \left(\frac{1}{4}\right)\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$ and $\gamma = 5$, then $P_F = e^{K_b} \cdot P_I$ produces final probabilities

$$\begin{bmatrix} p_{GA} = .08 & p_{GW} = .42 & p_{BA} = .42 & p_{BW} = .08 \end{bmatrix}.$$

For the quantum model, we set

$$H_b = H_{bA} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + H_{bW} \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

where

$$H_{bA} = \gamma \cdot \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad H_{bW} = \gamma \cdot \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

The submatrix H_{bA} rotates beliefs toward a 'bad' guy category when the decision maker tends toward an 'attack' action, and the other submatrix H_{bW} rotates beliefs toward a 'good' guy category when the decision maker tends toward a 'withdraw' action. For example, if $Q_I' = \left(\frac{1}{2}\right)\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$ and $\gamma = 1/\sqrt{2}$, then $Q_F = e^{-i\cdot H_b} \cdot Q_I$ produces final probabilities

$$\begin{bmatrix} |q_{GA}|^2 = .07 & |q_{GW}|^2 = .43 & |q_{BA}|^2 = .43 & |q_{BW}|^2 = .07 \end{bmatrix}.$$

But using K_b or H_b alone does not change the choice probabilities for the actions. Therefore, we need both factors, $K = K_b + K_d$ for the Markov model, and $H = H_b + H_d$ for the quantum model, to simultaneously change beliefs and actions.

Using $K = K_d + K_b$ for the Markov model, the probabilities for the D-Alone condition are related to the probabilities for the C-then-D condition by expressing the initial probabilities for the D-Alone condition in terms of a mixture of the initial probabilities conditioned on each categorization response:

$$\begin{aligned} L \cdot M_A \cdot P_F &= L \cdot M_A \cdot e^{(K_d+K_b)\cdot t} \cdot P_I \\ &= L \cdot M_A \cdot e^{(K_d+K_b)\cdot t} \cdot (p_G \cdot P_G + p_B \cdot P_B) \\ &= p_G \cdot L \cdot M_A \cdot e^{(K_d+K_b)\cdot t} \cdot P_G \\ &\quad + p_B \cdot L \cdot M_A \cdot e^{(K_d+K_b)\cdot t} \cdot P_B. \end{aligned} \tag{13}$$

One can see from the above equation that the 'attack' probability for D-Alone condition is again a weighted average of the probabilities conditioned on each category response. Therefore, as expected, including the dissonance effect does not help, and the Markov model still cannot explain violations of the law of total probability observed for the narrow faces. Using $H = H_d + H_b$ for the quantum model, the amplitudes for the D-Alone condition are related to the amplitudes for the C-then-D condition by expressing the initial amplitudes for the D-Alone condition in terms of a superposition of the amplitudes for the two categorized states:

$$\begin{aligned} M_A \cdot Q_F &= M_A \cdot e^{-i\cdot(H_d+H_b)\cdot t} \cdot Q_I \\ &= M_A \cdot e^{-i\cdot(H_d+H_b)\cdot t} \cdot (q_G \cdot Q_G + q_B \cdot Q_B) \\ &= q_G \cdot M_A \cdot e^{-i\cdot(H_d+H_b)\cdot t} \cdot Q_G + q_B \cdot M_A \cdot e^{-i\cdot(H_d+H_b)\cdot t} \cdot Q_B. \end{aligned} \tag{14}$$

One can see from the above equation that the amplitudes that determine the 'attack' decision for the D-Alone condition are a superposition of the amplitudes conditioned on each category response. However, the probabilities are obtained by squaring the magnitudes of the amplitudes corresponding to each action, and now the interference arises because $Q^{\dagger}_{A|G} \cdot Q_{A|B} \neq 0$. The dissonance effect causes these two vectors to become non-orthogonal.

Altogether, both of the four-dimensional models, Markov and quantum, require estimating four parameters $\{|q_G|^2 = p_G, \mu_G, \mu_B, \gamma\}$ from four data points $\{\Pr(G), \Pr(A|G), \Pr(A|B), \Pr(A)\}$. As pointed out earlier, there do not exist any parameters that will allow the Markov model to fit these data and violate the law of total probability. We could introduce new hidden states into the Markov model to do this. For example, one could assume that there is a fifth hidden state, $\langle \tilde{G} \wedge \tilde{B} \rangle$, which is only entered in the D-Alone condition. Then the law of total probability would fail because $\Pr(A)$ includes a contribution from this hidden state that does not appear in either $\Pr(A|B)$ or $\Pr(A|G)$. However, this five-dimensional Markov model is so complex that it can fit any pattern of data for the present experiment, and so we will not pursue that here.

The quantum model can fit the results, but in order to provide a one degree of freedom test, we need to constrain at least one parameter for each condition. For the wide face condition, we fixed

Author's personal copy

$\gamma = 0$ and used the estimates $|q_G|^2 = .82$, $\mu_G = .14$, $\mu_B = .03$. These predictions are shown in the bottom group of rows in Table 1.1, and they exactly reproduced the observed proportions for the average data processed in the C-then-D order. For the narrow face condition, we fixed μ_B and used the estimates $|q_G|^2 = .19$, $\mu_G = -.35$, $\gamma = .96$; these predictions are also shown in the bottom group of rows in Table 1.1. The predictions for the narrow faces are close but not perfect, yet they succeed in reproducing the basic pattern of the violation of the law of total probability observed under this condition. In summary, the Markov model can fit the data for the wide faces but it cannot fit the data for the narrow faces using any parameters; the quantum can fit both the wide and narrow face conditions, but not perfectly for the narrow faces, and this requires three parameters to fit four data points, which is not very impressive at this point. Furthermore, we do not have an explanation for the differences between narrow and wide face results. Clearly, more theoretical work and stronger experimental tests are needed.

3. Discussion

3.1. Empirical case for interference effects

Empirical violations of the law of total probability are often called interference effects. One of the main reasons for the invention of quantum theory by physicists was to explain interference effects observed in particle physics (Feynman & Hibbs, 1965). Now researchers have observed interference effects with human choices, which motivates the application of quantum probability to this domain (Khrennikov, 2004). What is the collection of evidence for these interference effects?

The first example of interference effects was obtained from studies using a two-stage gambling paradigm (Tversky & Shafir, 1992). Participants decided whether or not to play a gamble a second time after winning or losing the first time. When told that they won the first play, 69% chose to play again; when told that they lost the first play, 59% chose to play again; but when the first play was unknown only 36% chose to play again. The law of total probability implies that the unknown probability should be a weighted average of the two known probabilities, which is dramatically violated in this study. Tversky and Shafir (1992) found this result in two separate experiments: one using a within-subjects design where each person experienced all three conditions, and the other using a between-subjects design using a different group of people for each condition. More recently, however, these results were not replicated in a series of studies (Kuhberger, Komunska, & Perner, 2001).

The second example of interference effects were obtained using a prisoner dilemma game (Shafir & Tversky, 1992). Participants decided whether or not to defect depending on knowledge of an opponent's decision. When told that the opponent will defect, 97% chose to defect; when told that the opponent will cooperate, 84% chose to defect; but when the opponent's strategy was unknown, only 63% chose to defect. Once again, the law of total probability implies that the unknown probability should be a weighted average of the two known probabilities, which is clearly violated in this study. This finding was replicated by Busemeyer, Wang, and Townsend (2006), who found 91% defection when the opponent was known to defect, 84% defection when the opponent was known to cooperate, and only 66% defection for the unknown case. This pattern of results was also found, although not quite as strongly, in (Croson, 1999; Li & Taplin, 2002).

The third example of interference effects was obtained using a categorization–decision task in which participants were asked to categorize and/or choose an action. Townsend et al. (2000) found that one fourth of their participants produced statistically significant violations of the law of total probability. We replicated this experiment and found interference effects for the narrow face condition: if a face was categorized as a 'good' guy then it was 'attacked' in 43% of the trials; if a face was categorized as a 'bad' guy, then it was 'attacked' in 63% of the trials; but when no categorization was made, it was 'attacked' in 69% of the trials. Again, no weighted average of the two known cases is equal to the unknown case.

Finally, interference effects were recently found in a series of four perceptual judgment tasks with impossible figures (Conte et al., 2009). These experiments employed two conditions: In in condition A-Alone, a single binary choice was made between two interpretations of an impossible figure (labeled positive versus negative); in the other condition, call it B-then-A, a binary judgment about a different impossible figure B was made before the judgment about figure A. Here we briefly review results of the first of these studies, which in fact replicated the stimuli and results from an earlier study (Conte et al., 2006). If B was given a positive interpretation, then the proportion of positive A interpretations was .60; if B was given a negative interpretation, then the proportion of positive A interpretations was .375; but under the A-Alone condition, the proportion of positive A interpretations was .67. Furthermore, this difference was statistically significant according to a chi-square test ($p = .017$). In sum, there is growing empirical evidence for interference effects in a variety of decision making paradigms in psychology.

3.2. Empirical case for double stochasticity

Transition matrices describe the conditional probability of observing the next (row) state of a system conditioned on the previous (column) state. Transition matrices require that the rows within a column sum to one. Double stochasticity requires both the rows and columns of a transition matrix to sum to unity. Quantum models require double stochasticity but Markov models do not. So what is the empirical evidence regarding double stochasticity?

The answer depends on the dimensionality of the state space. Let us first consider the two-dimensional case, which is easier to test. In this case, transitions from belief states $B = \{|B_1\rangle, |B_2\rangle\}$ to action states $A = \{|A_1\rangle, |A_2\rangle\}$ produce a 2×2 transition matrix. Double stochasticity requires that $\Pr(A_1|B_i) + \Pr(A_2|B_i) = 1$ and that $\Pr(A_i|B_1) + \Pr(A_i|B_2) = 1$, and the latter constraint provides the critical test. In other words, the probability of an action must sum to one across the two known belief states.

For the two-stage gambling paradigm (Tversky & Shafir, 1992), letting A_1 stand for playing the gamble, and letting B_i stand for knowledge about the outcome of the first play, then the two critical probabilities were $\Pr(A_1|B_1) = .69$, and $\Pr(A_1|B_2) = .59$, which violate the critical property. For the prisoner dilemma paradigm (Shafir & Tversky, 1992), letting A_1 represent the player's decision to defect and letting B_i represent knowledge of the opponent's action, the two critical probabilities were $\Pr(A_1|B_1) = .97$, and $\Pr(A_1|B_2) = .84$, which also clearly violate this property. For the category–decision making experiment, the conditional probabilities (shown in Table 1.1) violate this property for the narrow faces. In sum, double stochasticity frequently fails for the two-dimensional transition matrix.

Rather than restricting transitions from states in B to states in A, we could allow transitions between combination states in $S = \{|G, A\rangle, |G, W\rangle, |B, A\rangle, |B, W\rangle\}$. This produces a 4×4 transition matrix, and the four state quantum model continues to obey the law of double stochasticity for this 4×4 transition matrix (defining transitions between states in the set S). Unfortunately, this is more difficult to empirically test, and none of the experiments provide a test of double stochasticity for the four-dimensional transition matrix. Future research is needed to test double stochasticity for the four-dimensional case.

Author's personal copy

3.3. Theoretical explanations

How can one explain the empirical findings of interference effects? One cannot immediately jump to the conclusion that all Markov models fail and a quantum model is necessary. There always remains a possibility of constructing a classic Markov model for these effects by adding new hidden states to the model. One cannot also jump to the conclusion that the quantum model is more complex. Adding new states to a Markov model can make it considerably more complex than a smaller-dimensional quantum model. In fact, Markov and quantum models simply obey different probability laws — the former obeys the law of total probability and the latter obeys the law of double stochasticity.

The best we can do is to compare a specific n-dimensional Markov model with a specific n-dimensional quantum model. But we would like to test these specific models in a parameter free manner by using properties such as the law of total probability or the law of double stochasticity that are defined with respect to the model dimension. So far, we have compared specific Markov and quantum models for the two-stage gambling paradigm and prisoner dilemma paradigm (Pothos & Busemeyer, 2009) and for the categorization–decision paradigm (Busemeyer & Wang, 2007). We established that a two-dimensional Markov model fails, because it must satisfy the law of total probability (which was violated); likewise a two-dimensional quantum model fails because it must satisfy the law of double stochasticity for a 2×2 transition matrix (which was also violated). We also established that the four-dimensional Markov model also fails because it continues to satisfy the law of total probability (which still fails); and we demonstrated that the four-dimensional quantum model can approximately fit the results from all three paradigms. Unfortunately we could not test double stochasticity for the four-dimensional quantum model, which makes the current comparison of the Markov and quantum four-dimensional models not exactly fair.

The four-dimensional quantum model generates the interference effect for all three paradigms: two-stage gambling, prisoner dilemma, and the categorization–decision, and it does so by using the same principle. The key idea is that we assume a Hamiltonian that contains two factors — one factor evolves preferences over time based on payoffs in the task, and the other factor evolves beliefs over time. Thus preferences and beliefs evolve simultaneously across time during deliberation leading to an action decision. Furthermore, we assume that beliefs and preferences tend to evolve toward consistency (a cognitive consistency effect). For example, in the prisoner dilemma game, if you tend toward defecting then you also tend to believe your opponent will defect; in the category–decision paradigm, if you tend toward 'attacking' then you also tend to believe the face is a 'bad' guy; in the two-stage gambling paradigm, if you intend to play the gamble a second time, then you tend to believe you are a winner. This two-factor Hamiltonian produces a unitary transformation that entangles beliefs and preferences over time, and this entangled state generates the interference term for the action probability. In sum, the interference depends on the parameter γ in the Hamiltonian of the quantum model. If this is set to zero, the interference disappears; if it is positive, it rotates beliefs in a manner that makes them consistent with preferences; if it is negative, it rotates beliefs to be inconsistent with preferences. In all three paradigms, the sign of this parameter was in the same direction of producing beliefs that were consistent with preferences (i.e., γ has the same direction of effect in all three paradigms). Finally, recall that this same principle, applied to the intensity matrix, does not help the Markov model, and it still satisfies the law of total probability. So we need both cognitive consistency and quantum probability to produce the interference effect.

While it is encouraging that the same four-dimensional quantum model can fit the interference effects from three quite different paradigms, the achievement is far from convincing yet for several reasons. First, the predictions derived from the quantum model require three parameters to fit four data points, and much stronger quantitative tests of the model are needed in future studies using a much higher ratio of data points to parameters. Second, for the categorization–decision paradigm, we have no explanation for why interference effects occur with the narrow faces, but not the wide faces. However, we think this is an important beginning that could guide future experimental tests of these models.

3.4. Psychology of Markov and quantum models

What is the psychological difference between Markov and quantum models? How does one interpret the parameters of these models? Markov models are already well established in psychology. For example, diffusion models of signal detection (Ratcliff & Smith, 2004), stochastic models of information processing (Townsend & Ashby, 1983), multinomial processing trees for memory (Batchelder & Reiffer, 1999), and random walk models of decision making (Busemeyer & Townsend, 1993) are all Markov models. What we have tried to show here and elsewhere (Busemeyer et al., 2006) is the great similarity between Markov and quantum models. By analogy, we argue that quantum models should be useful to psychologists as well.

Both Markov and quantum models assume an initial state that can be biased by prior information. Both Markov and quantum models have operators that evolve belief and preference states during deliberation before making a decision — the dynamics of the Markov model are derived from the Kolmogorov differential equation, and the dynamics for the quantum model are derived from the Schrödinger differential equation. Both the Markov and quantum models use the same kind of parameters determined by the strength of evidence and/or the utility of actions — the Markov model introduces these parameters into the intensity matrix of the Kolmogorov equation, and the quantum model introduces these into the Hamiltonian matrix of the Schrödinger equation. Both the Markov and quantum models derive the final choice probabilities from the states that evolved from deliberation. The key difference is the probability law used to derive these choice probabilities — the Markov model operates directly on probabilities and the whole process remains linear; the quantum model operates linearly on amplitudes but the response probabilities are determined by their squared magnitudes. The latter introduces a critical nonlinearity into the whole process. In short, the psychological interpretation of parameters is almost identical for the two models — the key difference comes from the difference in probability laws. The use of complex amplitudes in the quantum model allows for interference effects to arise from the squared magnitudes.

What are the psychological implications of these different probability laws? The usual interpretation of the Markov model is the following. It assumes that at any point in time during deliberation, a person is exactly in one of the basis states. We just do not know which one. The process steps from one basis state to another from moment to moment during deliberation. A single decision is the end result of a sequence of transitions that forms a single trajectory from the beginning to the final state. (Diffusion processes are continuous Markov processes derived by taking the limit of small discrete state changes made in small time increments.) A single trajectory is a deterministic function of time, similar to generating a sequence of random numbers from a fixed seed in a computer simulation. At the end of deliberation, and immediately before a decision is made, the person is in a determinant basis state. The choice made at the time of decision simply manifests this pre existing state. The Markov model derives the probability distribution over all the possible trajectories (e.g.,

Author's personal copy

assigns a probability to each seed in a computer simulation). There is only one trajectory for any given decision and so they cannot interfere with each other. In short, the Markov model assumes that the brain does its processing like a particle and obeys the probability laws for particle motion.

A way to interpret the quantum model is as follows. (There is no uncontroversial interpretation!) It assumes the person is *not* exactly in any basis state at each point in time during deliberation. Instead the person is in a superposed state at each moment in time. In this superposed state, all the states coexist in parallel at any moment, and this indeterminate or fuzzy state flows across time until the point when a decision must be made. Even for a single decision, one cannot assume that a single (albeit unknown) trajectory is followed. Instead, all of the indeterminate trajectories coexist in parallel across time, producing a wave of potentialities flowing over the states across time. Immediately before the final decision, the person is not in a clear and determinant basis state; and instead, the observed decision creates a clear and determinant basis state. The quantum model derives the probability amplitudes for all possible trajectories, and all of these indeterminate trajectories can interfere with one another. In short, the quantum model assumes the brain does its processing like a wave and obeys the laws for wave motion.

3.5. Related research

In addition to our own work, a number of other researchers have been developing quantum decision models for describing human decision making in different applications. Three of the earliest were models for interference effects with attitude questions (Aerts & Aerts, 1994), perceptual decisions (Atmanspacher, Filk & Romer, 2004), cognitive judgments (Khrennikov, 1999), and probability estimation (Bordley & Kadane, 1999). More recently, models of approach-avoidance behavior have been proposed (Ivancevic & Aidman, 2007). In this special issue, one can find applications of quantum models to preferential choice by Mogiliansky-Lambert, Zamir, and Zwirn, and to utility theory by La Mura, and to probability judgments by Franco and Aerts, and finally to financial arbitrage by Khrennikov and Haven.

Appendix A. Instructions to participants

Closely following Townsend et al. (2000), the scenario was set up using the following instruction for the categorization then decision (C-then-D) condition (and with small alternations for the other conditions at the beginning of an experiment session:

"You have been chosen by NASA to travel to the planet Meboo to find out more about two colonies, the Adoks and the Lorks.

As you interact with the two colonies, you will be first asked to categorize each face as either an 'Adok' or a 'Lork'. The Adoks tend to have round faces and thin lips, and the Lorks tend to have narrow faces with fat lips. But, this is not absolute! As in any culture, there is cross-over. A face with the features of an Adok may actually be a Lork, and a face with the features of a Lork may actually be an Adok.

You have up to 10 s to view each face (you may answer before the 10 s are up). You should press the key '1' (labeled 'A/F') for an 'Adok' or '2' (labeled 'L/D') for a 'Lork'.

Then, you have a choice to make: you can be friendly or defensive to the face. Adoks have the tendency to be friendly while Lorks tend to be hostile. This is not absolute! Since you do not know how the individual will act towards you, make your decision carefully.

You should press the key '1' (labeled as 'A/F') for Friendly or '2' (labeled as 'L/D') for Defensive. Again, you have up to 10 s to make the decision.

You will be given feedback for your categorization and action decision after each face. Then, click the spacebar (labeled "continue") to continue to the next face. There are 17 faces following this instruction".

Appendix B. Hilbert space representation

The two-dimensional quantum model can be represented as a Hilbert space model by using the following standard quantum theory definitions and relations. First, we assume a two-dimensional Hilbert space H_2. The two category states $\{|G\rangle, |B\rangle\}$ form a pair of orthonormal basis vectors for H_2, and the two action states $\{|A\rangle, |W\rangle\}$ form another pair of orthonormal basis vectors for H_2. The two bases are related by

$$|G\rangle = \langle A|G\rangle \cdot |A\rangle + \langle W|G\rangle \cdot |W\rangle ,$$
$$|B\rangle = \langle A|B\rangle \cdot |A\rangle + \langle W|B\rangle \cdot |W\rangle ,$$

where $\langle x|y\rangle$ is the Dirac notation for the inner product between two vectors in H. The matrix U is the transformation matrix that transforms coordinates of $\{|G\rangle, |B\rangle\}$ into coordinates of $\{|A\rangle , |W\rangle\}$:

$$U = \begin{bmatrix} U_{AG} & U_{AB} \\ U_{WG} & U_{WB} \end{bmatrix} = \begin{bmatrix} \langle A|G\rangle & \langle A|B\rangle \\ \langle W|G\rangle & \langle W|B\rangle \end{bmatrix}.$$

This matrix must be unitary for the following reasons. Note that

$$\begin{aligned} |\langle A|G\rangle|^2 + |\langle A|B\rangle|^2 &= \langle A|G\rangle \langle G|A\rangle + \langle A|B\rangle \langle B|A\rangle \\ &= \langle A|(|G\rangle \langle G| + |B\rangle \langle B|)|A\rangle = \langle A|I|A\rangle \\ &= 1, \end{aligned}$$

and similarly $|\langle W|G\rangle|^2 + |\langle W|B\rangle|^2 = 1$. Also note that

$$\begin{aligned} |\langle A|G\rangle|^2 + |\langle W|G\rangle|^2 &= \langle G|A\rangle \langle A|G\rangle + \langle G|W\rangle \langle W|G\rangle \\ &= \langle G|(|A\rangle \langle A| + |W\rangle \langle W|)|G\rangle = \langle G|I|G\rangle = 1, \end{aligned}$$

and similarly $|\langle A|B\rangle|^2 + |\langle W|B\rangle|^2 = 1$. Thus both rows and columns must be unit length. Finally, note that

$$\begin{aligned} \langle G|B\rangle &= 0 = (\langle G|A\rangle \langle A| + \langle G|W\rangle \langle W|) \cdot (\langle A|B\rangle |A\rangle + \langle W|B\rangle |W\rangle) \\ &= \langle G|A\rangle \langle A|B\rangle + \langle G|W\rangle \langle W|B\rangle , \end{aligned}$$

and so the columns of U must be orthogonal. The state of the quantum system is represented by a unit length vector in H:

$$\begin{aligned} |q\rangle &= q_G \cdot |G\rangle + q_B \cdot |B\rangle \\ &= q_G \cdot [\langle A|G\rangle \cdot |A\rangle + \langle W|G\rangle \cdot |W\rangle] + q_B \cdot [\langle A|B\rangle \cdot |A\rangle \\ &\quad + \langle W|B\rangle \cdot |W\rangle] \\ &= (q_G \cdot U_{AG} + q_B \cdot U_{AB}) \cdot |A\rangle + (q_G \cdot U_{WG} + q_B \cdot U_{WB}) \cdot |W\rangle . \end{aligned}$$

For the C-then-D condition, the probability of observing each category response is obtained by the squared projections $\widehat{\Pr}(G) = |\langle G|q\rangle|^2 = |q_G|^2$ and $\widehat{\Pr}(B) = |\langle B|q\rangle|^2 = |q_B|^2$. After observing a category decision, the state changes to either $|G\rangle$ following a 'good' guy categorization or $|B\rangle$ following a 'bad' guy categorization. The probability of observing an 'attack' action after observing each category response is obtained by the squared projections $\widehat{\Pr}(A|G) = |\langle A|G\rangle|^2 = |U_{AG}|^2$ and $\widehat{\Pr}(A|B) = |\langle A|B\rangle|^2 = |U_{AB}|^2$. For the D-Alone condition, the probability of taking an 'attack' action equals $\widehat{\Pr}(A) = |\langle A|q\rangle|^2 = |(q_G \cdot U_{AG} + q_B \cdot U_{AB})|^2$. These are the same probabilities that we obtained from the two-dimensional quantum model expressed as probability amplitudes.

For the four-dimensional quantum model we hypothesize a four-dimensional Hilbert space H_4. The set of four orthonormal basis vectors $S = \{|GA\rangle , |GW\rangle , |BA\rangle , |BW\rangle\}$ forms an orthonormal basis for H_4. The initial state is a superposition $|P_I\rangle = q_{GA} \cdot |GA\rangle + q_{GW} \cdot |GW\rangle + q_{BA} \cdot |BA\rangle + q_{BW} \cdot |BW\rangle$. Thus the initial amplitude vector, P_I, described earlier, contains the coordinates of the initial state $|P_I\rangle$ with respect to the S basis. The initial state is transformed into the final state by a unitary operator: $|P_F\rangle = \mathbf{U}(t) \cdot |P_I\rangle$. This operator must be unitary to preserve inner products, which is required for this Hilbert space representation. The matrix rep-

Author's personal copy

resentation of this unitary operator with respect to the S basis is $U(t) = [U_{ij}(t)] = [\langle i|\mathbf{U}(t)|j\rangle, |i\rangle, |j\rangle \in S]$, and this is determined by the Schrödinger equation (8). The coordinates of the final state vector $|P_F\rangle$ with respect to the S basis are given by the final amplitude vector $P_F = U(t) \cdot P_I$. The measurement operator for the 'attack' action is defined by the projector $\mathbf{M}_A = |GA\rangle\langle GA| + |BA\rangle\langle BA|$. The squared projection $|\mathbf{M}_A \cdot |P_F\rangle|^2$ equals the probability of observing an 'attack' action. The matrix representation of this projector with respect to the S basis is $M_A = [\langle i|\mathbf{M}_A|j\rangle, |i\rangle, |j\rangle \in S]$, which, as described earlier, is a diagonal matrix with ones in the first and third rows and zeros elsewhere.

References

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

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.

Atmanspacher, H., Filk, T., & Romer, H. (2004). Quantum Zeno features of bistable perception. *Biological Cybernetics*, *90*, 33–40.

Batchelder, W. H., & Reiffer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review*, *6*, 57–86.

Bhattacharya, R. N., & Waymire, E. C. (1990). *Stochastic processes with applications*. Wiley.

Bordley, R., & Kadane, J. B. (1999). Experiment-dependent priors in psychology. *Theory and Decision*, *47*(3), 213–227.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognitive approach to decision making. *Psychological Review*, *100*, 432–459.

Busemeyer, J. R., & Wang, Z. (2007). Quantum information processing explanations for interactions between inferences and decisions. In P. Bruza, W. Lawless, K. van Rijsbergen, & D. A. Sofge (Eds.), *Quantum interaction: Papers from the AAAI spring symposium* (pp. 91–97). AAAI Press.

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

Conte, E., Khrennikov, A. Y., Todarello, O., Federici, A., Mendolicchio, L., & Zbilut, J. P. (2009). Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Systems and Information Dynamics*, 1–17.

Conte, E., Todarello, O., Federici, A., Vitiello, F., Lopane, M., & Khrennikov, A. (2006). 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, and Fractals*, *31*, 1076–1088.

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

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Feynman, R., & Hibbs, A. (1965). *Quantum mechanics and path integrals*. McGraw-Hill.

Gudder, S. P. (1988). *Quantum probability*. Academic Press.

Ivancevic, V., & Aidman, E. (2007). Life space foam: A medium for motivational and cognitive dynamics. *Physica A*, *382*, 616–630.

Kayser, A. (1985). *Heads*. Abbeville Press.

Khrennikov, A. Y. (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. Y. (2004). *Information dynamics in cognitive, psychological, social and anomalous phenomena*. Kluwer Academic.

Khrennikov, A. Y. (2007). Can quantum information be processed by macroscopic systems. *Quantum Information Processing*, *6*(6), 401–429.

Kuhberger, A., Komunska, D., & Perner, J. (2001). The disjunction effect: Does it exist for two-step gambles? *Organizational Behavior and Human Decision Processes*, *85*(2), 250–264.

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

Peres, A. (1998). *Quantum theory: Concepts and methods*. Kluwer Academic.

Pothos, E. H., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of 'rational' decision theory. *Proceedings of the Royal Society B*.

Ratcliff, R., & Smith, P. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.

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

Townsend, J. T., & Ashby, G. F. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge University Press.

Townsend, J. T., Silva, K. M., Spencer-Smith, J., & Wenger, M. (2000). Exploring the relations between categorization and decision making with regard to realistic face stimuli. *Pragmatics and Cognition*, *8*, 83–105.

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