Cognition 150 (2016) 133–149

Contents lists available at ScienceDirect

Cognition

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

Interference effects of categorization on decision making ☆

Zheng Wang [a,*], Jerome R. Busemeyer [b]

[a] School of Communication, The Ohio State University, 3145 Derby Hall, 154 N. Oval, Columbus, OH 43210, United States
[b] Indiana University, United States

ARTICLE INFO

Article history:
Received 20 October 2014
Revised 26 January 2016
Accepted 28 January 2016

Keywords:
Categorization
Decision
Interference effects
Law of total probability
Entanglement
Superposition
Quantum probability
Markov model
Signal detection model

ABSTRACT

Many decision making tasks in life involve a categorization process, but the effects of categorization on subsequent decision making has rarely been studied. This issue was explored in three experiments ($N = 721$), in which participants were shown a face stimulus on each trial and performed variations of categorization-decision tasks. On C-D trials, they categorized the stimulus and then made an action decision; on X-D trials, they were told the category and then made an action decision; on D-alone trials, they only made an action decision. An interference effect emerged in some of the conditions, such that the probability of an action on the D-alone trials (i.e., when there was no explicit categorization before the decision) differed from the total probability of the same action on the C-D or X-D trials (i.e., when there was explicit categorization before the decision). Interference effects are important because they indicate a violation of the classical law of total probability, which is assumed by many cognitive models. Across all three experiments, a complex pattern of interference effects systematically occurred for different types of stimuli and for different types of categorization-decision tasks. These interference effects present a challenge for traditional cognitive models, such as Markov and signal detection models, but a quantum cognition model, called the belief-action entanglement (BAE) model, predicted that these results could occur. The BAE model employs the quantum principles of superposition and entanglement to explain the psychological mechanisms underlying the puzzling interference effects. The model can be applied to many important and practical categorization-decision situations in life.

1. Introduction

The fields of categorization and decision making are empirically mature and theoretically well developed, but to a large degree, they have evolved in a parallel and independent manner. Little is known about the interactions between these two basic cognitive tasks – that is, how a categorization task changes performance on a subsequent decision task.[1] In many situations in life, decision makers need to make categorizations before deciding on an action. For example, a doctor needs to categorize a biopsy as cancerous or not before making treatment decisions; a judge needs to categorize a defendant as guilty or not before assigning a punishment; a police officer needs to categorize a driver as intoxicated or not before making an arrest; a military operator needs to categorize an agent as an enemy or not before making an attacking decision. In all these examples, it seems necessary to infer a category before choosing an action. Suppose the decision maker has to *report* this category inference before making the decision. How does this overt report of the category affect the later decision? For example, would the probability that a police officer shoots a suspect be changed if she or he had to report seeing a weapon possessed by the suspect first?

In the work described below, participants were presented with a face and were asked to categorize it first and then decide on an action. However, the general categorization-decision paradigm is not limited to these particular details, and as mentioned above, there are many important and practical examples of categorization-decision situations in real life. In general, any task that has the following four characteristics falls into this paradigm: (1) a stimulus providing information is presented, after which (2) a categorical inference is made based on the stimulus, followed by (3) a decision about an action, and (4) the action has consequences that depend on both the action and the true state of the category.

To explore the relation among these tasks, three experiments were conducted, and three theoretical explanations – a Markov

☆ This research was supported by the US NSF SES-1153726, SES-1153846, the US AFOSR FA 9550-12-1-0397, and FA 9550-15-1-0343.

* Corresponding author.
E-mail addresses: wang.1243@osu.edu, zhengjoycewang@gmail.com (Z. Wang).

[1] Maddox and Bohil (1998) examined the effects of decision making variables such as prior probabilities and payoffs on a categorization task, but here we examine how a categorization task affects a subsequent decision task. More closely related is the effect of categorization on subsequent feature inferences, such as has been discussed by Murphy and Ross (1994), Griffiths, Hayes, and Newell (2012), and Chaigneau, Barsalou, and Sloman (2004). In the General Discussion section, we relate our research to these other lines of work.

http://dx.doi.org/10.1016/j.cognition.2016.01.019
0010-0277/

model, a signal detection model, and a quantum cognition model based on quantum probability rules – are discussed and compared. Only the quantum cognition model *a priori* predicted an interference effect of categorization on subsequent decision making that systematically occurred in the experiments.

2. The categorization-decision paradigm

2.1. The categorization-decision interference

Townsend, Silva, Spencer-Smith, and Wenger (2000) initiated an investigation of the category-decision paradigm. On each trial, participants were shown one of 34 faces that were assigned to a "good guy" or "bad guy" category based on some facial features (e.g., width of faces), and then asked to decide whether to "attack" the face or "withdraw" from it. Fig. 1 illustrates some examples of the faces used in our new experiments, which were similar to those employed by Townsend et al. As shown, it was fairly easy to discriminate the two types of faces, but the task was made difficult because the assignment of faces to a category was probabilistic: The narrow faces were assigned to the "bad guy" category on 60% of the trials and to the "good guy" category on the remaining trials; likewise, the wide faces were assigned to the "good guy" category on 60% of the trials and to the "bad guy" category on the remaining trials.

The category was important because participants were rewarded on 70% of the trials for attacking faces that were assigned to the bad guy category and punished on 70% of the trials for attacking faces that were assigned to the good guy category. Likewise, they were rewarded on 70% of the trials for withdrawing from faces assigned to the good guy category and punished on 70% of the trials for withdrawing from faces assigned to the bad guy category. Participants were given six blocks of training, during which they first categorized a face and then decided on an action, and afterwards feedback was provided on both the category and the decision. The key manipulation occurred during a transfer test phase, during which each person received two additional blocks with three types of trials: (1) categorization and then decision (C-D) trials exactly like the original training, (2) categorization (C-alone) trials in which only a categorization was made with feedback, and (3) decision (D-alone) trials in which only a decision was made with feedback. For example, on a D-alone trial, the person was shown a face, simply decided to attack or withdraw, and received feedback on the decision. Of course, the categorization of the face on the D-alone trial remained highly relevant to the action decision, and it seems some implicit inference about the category was necessary before participants made the decision even though they did not have to explicitly report this inference.

Using this paradigm, one can examine within each participant how the overt report of the category interferes with the subsequent decision by comparing the probability of attacking on the D-alone trials (denoted as $p(A)$ for a face type) with the total probability of attacking on the C-D trials (denoted as $p_T(A)$ for the same face type). The latter is simply the probability of attacking on C-D trials pooled across trials when the categorization response is ignored. It can also be expressed using the classical law of total probability, which states that the probability to attack (A) equals the probability that the person categorizes a face as a good guy (G) and then attacks plus the probability that the person categorizes the face as a bad guy (B) and then attacks: $p_T(A) = p(G \cap A) + p(B \cap A)$. If these two ways of determining the probability of attacking on D-alone and C-D trials agree for a participant, $p(A) = p_T(A)$, then we say that the law of total probability is empirically satisfied. Based on a chi-square test, Townsend et al. (2000) found that 25% of 138 participants produced statistically significant violations of this law. Apparently, the seemingly innocuous overt report of a category changed how a subsequent decision was made. Specifically, we define an *interference effect* of categorization on decision making as the difference between the probabilities of an event when it is measured alone versus when it is measured after another event, such as, in our context, the probability of attacking on the D-alone trials and the total probability of attacking pooled across the C-D trials.

Busemeyer, Wang, and Lambert-Mogiliansky (2009) further investigated this paradigm and discovered a more surprising result. Their study involved 26 participants, and each participant received both C-D trials and D-alone trials. As shown in the first two rows of Table 1, when a face was most frequently assigned to the good guy category (we denote this type of face as type g faces), there was almost no interference effect. However, when a face was most frequently assigned to the bad guy category (we denote this type of face as type b faces), the probability of attacking was significantly greater for the D-alone condition as compared to the C-D condition, violating the law of total probability ($p(A) > p_T(A)$ for type b faces). More surprisingly, the probability of attacking in the D-alone condition, which left the good or bad guy categorization unresolved, was even greater than the probability of attacking given that the person had already categorized the face as a bad guy in the C-D condition ($p(A) > p(A|B)$) for type b faces! It is surprising that for some reason, the overt categorization response interfered with the action decision by reducing the tendency to attack faces that most likely belonged to the bad guy category.

2.2. Candidate models for the categorization-decision paradigm

There are several models that can be considered for the application to the *general* categorization-decision paradigm (not just the particular example used in the current study). Below we briefly summarize five candidates. The first two, the optimal and probability matching models, are oversimplified but provide useful baselines for considering competing models for the paradigm. They predict no interference effects. The next two, Markov and signal detection models, are more general cognitive models, but they fail to predict any interference effects either in an *a priori* manner. The last is a quantum cognition model, which *a priori* predicts that an interference effect could occur.

2.2.1. Optimal model

The optimal model describes the optimal behaviors. According to the optimal model, the decision to attack should depend only on the face. If a type b face is presented, then it is always optimal to attack, and if a type g face is presented, then it is always optimal to withdraw. This follows from the fact that the probability of reward for attacking equals the probability that the type of face is assigned to the bad guy category (.60) times the probability that a reward is given for attacking a bad guy (.70), plus the probability that the same type of face is assigned to the good guy category (.40) times the probability that a reward is given for attacking a good guy (.30). That is, for a type b face, the total probability of being awarded for attacking equals $.60 \cdot .70 + .40 \cdot .30 = .54$, and the probability of reward for withdrawing is $1 - .54 = .46$, so the optimal model predicts that participants should always decide to attack when a type b face is presented. Likewise, the optimal model predicts that the participant should always decide to withdraw when a type g face is presented. These predictions hold regardless of whether the trial is a C-D trial or a D-alone trial, because the categorization response provides no new information for making the action decision. Therefore, the optimal model predicts no interference effect for the categorization-decision paradigm.

Fig. 1. Examples of two narrow faces (left pair) and two wide faces (right pair).

Table 1
Average choice probability results from Busemeyer et al. (2009) and Experiment 1.

Exp	Face	N	$p'(G)$	$p(G)$	$p(A\|G)$	$p(B)$	$p(A\|B)$	$P_T(A)$	$p(A)$	Int
2009	*b*	26	.23	.19	.43	.81	.64	.60	.69	.09
2009	*g*	26	.79	.83	.36	.17	.53	.38	.39	.01
Exp 1	*b*	126	.20	.21	.41	.79	.58	.55	.59	.04
Exp 1	*g*	138	.79	.78	.39	.22	.52	.41	.42	.01
Q Model	*b*	–	.20	.20	.37	.80	.61	.56	.62	.06
Q Model	*g*	–	.80	.80	.38	.20	.62	.43	.43	0

Note: First two rows are from Busemeyer et al. (2009) and the second two rows are from Experiment 1 of the current article. The last two rows are predictions from a quantum model. The symbols g, b refer to the type of face stimulus, the symbols G, B refer to the two categories, and A refers to the attacking action. $p(A)$ is estimated from the D-alone condition, and $p_T(A)$ is the total probability from the C-D condition. $p'(G)$ is the estimate from the C-alone condition, and $p(G)$ is from the C-D condition. The empirical results shown in this table were obtained by first obtaining estimates for each individual, and then averaging the estimates across all participants. The first two rows differ slightly from those of Table 1.1 in Busemeyer et al. (2009), who used estimates pooled across all trials and all individuals.

2.2.2. Probability matching model

According to the probability matching model, a person acts according to the exact probabilities involved at each stage of the paradigm. For example, if a type *b* face is presented on a C-D trial, then the probability that the person categorizes the face as bad equals .60 and the probability that it is categorized as good equals .40. If the face is categorized as bad, then the person attacks with a probability of .70, and if it is categorized as good, then the person attacks with a probability of .30. Therefore, the probability of attacking for the C-D condition equals $.60 \cdot .70 + .40 \cdot .30 = .54$, and this is also the probability to attack under the D-alone condition. Therefore, the probability matching model predicts no interference effect. According to the results from the C-D trials shown in Table 1, the choice probabilities deviated from probability matching. The probability of categorization is more extreme than probability matching, and the probability of taking each action is less extreme than probability matching.

2.2.3. Markov model

Townsend et al. (2000) initially proposed a simple Markov model for the category-decision task, which can be viewed as a generalization of the probability matching model. The central assumption of the Markov model is that the categorization depends on the face, but the action decision depends only on the categorization (and not on the face anymore). When a type of face (*b* or *g*) is presented, the person initially starts in either a good guy state (G) with probability $\phi(A)$ or in a bad guy state (B) with probability $\phi(B)$. From state G, the person can transit to the attack state (A) with probability $\phi(A|G)$; from state B, the person can transit to the attack state with probability $\phi(A|B)$. Likewise, from state G, the person can transit to the withdraw state (W) with probability $\phi(W|G)$; from state B, the person can transit to the withdraw state with probability $\phi(W|B)$. Then, the probability to categorize a face as a good guy and decide to attack on C-D trials equals the product of the transition probabilities, $\phi(G) \cdot \phi(A|G)$; the probability to categorize a face as a bad guy and decide to attack on C-D trials equals the product of the transition probabilities, $\phi(B) \cdot \phi(A|B)$. The probability of the attack decision on D-alone trials equals the probability of reaching a final state A by two different paths, which equals the sum of the path probabilities: $\phi(A) = \phi(G) \cdot \phi(A|G) + \phi(B) \cdot \phi(A|B)$. The latter shows that the Markov model is consistent with the law of total probability, and thus cannot account for the observed interference effect. Furthermore, according to the Markov model, $\phi(A|B)$ should be the same for both types of faces, *g* and *b*, which is contrary to the empirical findings (see Table 1). Later in the paper, we will evaluate a more general version of the Markov model that allows $\phi(A|B)$ to change across face types. However, as proved in Appendix A, as long as we assume that the model parameters do not change across C-D and D-alone trials, then *all* Markov models for this task must satisfy the law of total probability and fail to predict interference effects.

2.2.4. Signal detection model

The multi-dimensional signal detection model (e.g., Ashby & Townsend, 1986) is a generalization of the optimal model. The central idea of the signal detection model is that both the categorization and the decision depend only on the face itself. Faces are represented as points in a multi-dimensional face space Ω. For the categorization task, the face space is divided into two mutually exclusive and exhaustive category regions, R_G for the good guy category and R_B for the bad guy category ($R_G \cap R_B = \emptyset, R_G \cup R_B = \Omega$). When a particular face f is sampled on a trial, the categorization is determined by whether it falls into the R_G or R_B region. Likewise, for the decision task, the face space is divided into two mutually exclusive and exhaustive decision regions, R_A for the attack decision and R_W for the withdraw decision. When a particular face f is sampled on a trial, the decision is determined by whether it falls into the R_A or R_W region. A combination of category and action on a C-D trial is determined by the intersection of regions. For example, if $f \in R_B \cap R_A$, then the person categorizes the face as bad and decides to attack. Unlike in the Markov model, in the signal detection model, the probability of taking an action, conditioned on the categorization, still depends on the type of face. That is, $p(A|B)$ for type *b* faces does not equal $p(A|B)$ for type *g* faces. However, the total probability of the action decision does not depend on whether or not the person made a categorization because $p(f \in R_A)$ is predicted to be the same for both the D-alone condition and the C-D condition: For a given type of face, $p(A) = p(f \in R_A) = p(f \in (R_A \cap R_G) \cup f \in (R_A \cap R_B)) = p(f \in R_G) \cdot p(f \in R_A | f \in R_G) + p(f \in R_B) \cdot p(f \in R_A | f \in R_B) = p_T(A)$. Therefore, this model cannot account for the observed

interference effect either. Note that this is a general prediction of all signal detection models: It does *not* depend on the number of dimensions of the space (e.g., one or two dimensions), it does *not* depend on stimulus distribution assumptions (e.g., multivariate normal), and it does *not* depend on assumptions about the form of the boundaries (e.g., linear vs. quadratic). The above argument assumes only that the decision region does not change across C-D and D-alone types of trials. Later in this paper, we will consider a more relaxed version of the signal detection model that allows the decision boundaries to change across C-D and D-alone tasks.

2.2.5. Quantum model

Pothos and Busemeyer (2009) developed a quantum decision model, called the belief-action entanglement (BAE) model, to account for violations of the law of total probability obtained in a different task, a prisoner's dilemma decision task. Based on this model, Busemeyer et al. (2009) predicted and found an interference effect using the categorization-decision task described earlier. We briefly describe a simple version of the quantum model here, and later in the article we present the more general version of the model.[2]

The simplified quantum model is very similar to the simple Markov model described above. When a face –type *b* or type *g*– is presented, there is a potential to make either a good guy or a bad guy category response. The potential to categorize the face as a good guy is determined by an amplitude $\psi(G)$, and if this response is obtained, then it would produce a transition to state *G*. The potential to categorize the face as a bad guy is determined by an amplitude $\psi(B)$, and if that response is obtained, then it would produce a transition to state *B*. If the person is in state *G*, there is a potential to make the attack action (*A*) with amplitude $\psi(A|G)$; if the person is in state *B*, there is a potential to make the attack action with amplitude $\psi(A|B)$. Thus, on C-D trials, the amplitude for categorizing a face as good and then deciding to attack equals the product of the transition amplitudes, $\psi(G)\cdot\psi(A|G)$. The amplitude for categorizing a face as bad and then deciding to attack equals the product of the transition amplitudes, $\psi(B)\cdot\psi(A|B)$. On D-alone trials, the potential to make the attack decision equals the amplitude of reaching a final state *A* by two different paths, which equals the sum of the path amplitudes: $\psi(A)=\psi(G)\cdot\psi(A|G)+\psi(B)\cdot\psi(A|B)$. This all seems very similar to the Markov model but described by amplitudes instead of probabilities based on quantum theory.

In quantum theory, probabilities are obtained by squaring the magnitudes of the amplitudes. Thus, on C-D trials, the probability to categorize a face as good and then attack equals $|\psi(G)\cdot\psi(A|G)|^2$, and the probability to categorize it as bad and then attack equals $|\psi(B)\cdot\psi(A|B)|^2$; the total probability to attack equals $|\psi(G)\cdot\psi(A|G)|^2+|\psi(B)\cdot\psi(A|B)|^2$. In comparison, on D-alone trials, the probability to attack equals $|\psi(G)\cdot\psi(A|G)+\psi(B)\cdot\psi(A|B)|^2$, which equals the total probability from C-D trials plus a cross-product term called the *interference* term, which can be positive, negative, or zero. Therefore, the quantum model predicts that interference effects can occur. However, this simple version does not explain why the interference effect occurs only with type *b* faces but not type *g* faces. Furthermore, like the Markov model, this simple version of a quantum model predicts that $\psi(A|B)$ should be the same for both types of faces. Later, we present a more general model that allows $\psi(A|B)$ to change across face types, and this more general model can also account for the interaction of interference with types of faces.

[2] Here we only provide a brief description of the quantum model. The last two rows of Table 1 were computed using a more general quantum model. In Section 7, we describe the difference in psychological assumptions between the quantum and Markov models.

3. Experiment 1

The Busemeyer et al. (2009) experiment was based on a relatively small number ($N = 26$) of participants who completed a large number of training trials (six training blocks plus two transfer phase blocks, with 34 trials per block). In Experiment 1, we replicated and extended these initial results with variations on the original paradigm, which now used briefer training, and examined the robustness of the results. The large sample size also allowed us to examine the distribution of interference effects and the correlation between interference effects obtained with each type of face.

One way to account for the findings based on the signal detection model is to assume that the decision boundaries are contracted for C-D trials as compared to D-alone trials, which would produce a positive interference effect. The problem with this account is that it predicts a positive interference effect for both types of faces, but we find the effect only for type *b* faces and not for type *g* faces. So, this does not provide a very coherent account of the effects. Nevertheless, we can also test this hypothesis by examining the correlation between interference effects. If it is assumed that participants contract the "attack" boundary following a categorization on C-D trials, then we should find a positive correlation for interference effects between *b* and *g* faces across a large sample of participants.

3.1. Method

3.1.1. Participants

The participants were 169 undergraduate students recruited from a U.S. Midwest university for course extra credit. Of the participants, 58.58% were female; 87.57% were Caucasian, 5.33% were African American, 4.73% were Asian, and the rest identified themselves as "mixed" or "other." The average age was 20.56 ($SD = 1.02$).

3.1.2. Face stimuli

The set of face stimuli created by Busemeyer et al. (2009) was used. It included 34 head-shots of Caucasian men with a neutral facial expression. The stimuli were digitally altered to manipulate two salient cues: the shape of the face and the thickness of the lips. Half of the faces were narrow with thick lips, and the other half were round with thin lips. As shown by the examples in Fig. 1, the different types of faces were easy to discriminate. The facial cues were probabilistically related to the good versus bad guy category, which was fully disclosed to participants using the cover story described below. Specifically, the round faces with thin lips had a 60% chance to be assigned to the "Adok" (good guy) category and 40% to the "Lork" (bad guy) category; the narrow faces with thick lips had a 40% chance and a 60% chance, respectively. Then, the good guys had a 70% chance to be rewarded for a withdraw action and 30% for an attack action; the bad guys had a 30% chance and a 70% chance, respectively. For example, of the 17 round face stimuli in each block, a randomly selected 60% were assigned to the Adok (good guy) category, and of those Adok faces in each block, a randomly selected 70% were assigned to be friendly (i.e., to reward for a withdraw response).

3.1.3. Experimental procedure

Experiment 1 was similar to Busemeyer et al. (2009) except that we shortened the number of training blocks to enable a larger sample size of participants. (Few learning effects occurred because the explicit cover story instructions provided sufficient task

information.) In addition, the experiment consisted of two conditions, and the association between the face features and the types of faces was manipulated as a between-subjects factor in one of the conditions. For all blocks, the faces presented within each block were randomized across trials for each participant.

In the first condition, 61 participants completed three blocks of trials during a single session. Blocks 1 and 2 each included 34 C-D trials. Blocks 3 included 34 C-alone trials and 34 D-alone trials. Different from Busemeyer et al. (2009) in which the C-alone and D-alone trials were mixed together, this condition presented one block of C-alone trials, and another block of D-alone trials, with the order of blocks randomized across participants. The second block of the C-D trials is compared with the transfer phase (Block 3) in the report below.

In the second condition, 108 different participants completed three blocks of trials during a single session as well. This condition closely replicated the first condition, but added two changes. First, the C-alone and D-alone trials during Block 3 were mixed together and randomized. Second, we counterbalanced the assignment of face features (narrow, wide) and types (good, bad). Participants were randomly assigned to one of the associations between face features and types of faces: (1) the narrow faces with thick lips were more likely to be bad guys (the type *b* faces) or (2) they were more likely to be good guys (the type *g* faces).

Participants completed the experiment in groups of 2–10 using individual desktop computers. The task scenario was set up using the instructions and cover story employed by Townsend et al. (2000) and Busemeyer et al. (2009). At the beginning of the experiments, for the condition in which the narrow faces were assigned to be the type *b* faces, participants were told a story like the following: "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 thick 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 space bar (labeled "continue") to continue to the next face." For the condition in which the narrow faces were assigned to be the type *g* faces, the above cover story was modified to reflect the manipulation.

After reading the cover story, the participant viewed a series of faces on the computer. During each C-D trial, after a face stimulus was presented for 10 s, the participant was asked to categorize the face as Adok or Lork. Upon the categorization response, the participant was asked to select an action decision: to attack or to withdraw. Then, upon the decision response, feedback on both the categorization and decision was presented on the same screen for 3 s. For an Adok categorization response, if the face was pre-assigned as an Adok, the feedback would be "Yes! It was an Adok." If the face was pre-assigned as a Lork, it would be "No! It was not an Adok, but a Lork." For a Lork response, the feedback followed the same logic and format. For a withdraw response, if the Adok was pre-assigned to be friendly, 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, the feedback would say: "No! You were friendly to a hostile Adok. You were mugged." In similar ways, feedback was given to other response combinations. To facilitate the processing of the feedback information, pictures illustrating the action decision consequences (i.e., 20 dollars, a person being mugged) were presented on the feedback screen. For both the categorization and decision questions, the participant had up to 10 s to make a response using the assigned keys on the keyboard. If the participant failed to click either of the assigned keys within 10 s, a window popped up saying that "The time limit for this question has passed." Missing data were recorded. For each trial, after the feedback was presented at the end, 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.

The D-alone trials followed similar procedures. The only differences were that the participant was asked to make the action decision immediately after viewing the face; accordingly, feedback was given only on the decision, and the feedback lasted only 2 s. In Experiment 1, we also included C-alone trials, during which the participant was asked to make a categorization immediately after viewing the face; accordingly, feedback was given only on the categorization, and the feedback lasted only 2 s.

The pairing of narrow faces with the "bad" guy category produced a slightly larger interference effect; also, randomizing the C-alone and D-alone trials during the transfer test produced a slightly larger interference effect than the blocked procedure. However, these effects were small, and the pattern of interference effects was the same; therefore, we pooled the data across these conditions for presentation of the results.

3.2. Results

The estimated choice probabilities (i.e., sample proportions) were obtained for each participant and each type of face from the last block of C-D trials and from the transfer tests (D-alone trials, C-alone trials). Each estimate of a marginal probability is based on 17 choice trials per participant and each type of face: $p(G)$ and $p(B)$ denote the proportions for categorizing a face as good and bad, respectively, on C-D trials; $p_T(A)$ is the total proportion of attack choices across all C-D trials (combining the proportions through the two category selection paths); and $p(A)$ is the proportion of attack choices on D-alone trials. The difference $p(A) - p_T(A)$, computed for each person, defines the observed interference effect. Using the C-D trials, we also computed estimates of the conditional probabilities: $p(A|G)$ is the proportion choosing to attack given the face was categorized as good on C-D trials, and $p(A|B)$ is the proportion choosing to attack given the face was categorized as bad on C-D tests. We also obtained proportions for categorizing a face as good on transfer tests under a C-alone condition, which is denoted as $p'(G)$.

Some participants, whom we call "optimizers," always chose the "optimal" category for a particularly type of face on C-D trials: 43 did so for the narrow faces and 31 did so for the wide faces (approximately 25% and 18%, respectively, of the 169 participants). These participants obey the law of total probability for either type of face for trivial reasons, and for these participants, we cannot estimate the conditional probabilities for non-chosen categories and thus cannot really estimate the total probability for an action decision.

The second two rows of Table 1 shows the averages across non-optimizers for each type of face. As shown in the table, for the type *b* faces, a positive interference effect occurred (see the last column, labeled *Int*). However, for the type *g* faces, there was only a very small positive interference effect. It is also interesting that there

was no difference in the categorization results for C-D trials as compared to C-only trials, $p(G) \approx p'(G)$.

A statistical test was performed, separately for each type of face, using the interference effect obtained from each participant as the dependent variable. The statistical tests were computed using all 169 participants. The mean interference effect for the type b faces was statistically significant from zero ($t(168) = 2.24$, $SE = .015, p = .027$), but it was not significant for the g faces ($t(168) = .61, SE = .013, p = .54$).

There were strong correlations between $p(G)$ and $p'(G)$ ($r = .52, p < .0001$ for type b faces, $r = .65, p < .0001$ for type g faces). The correlations between $p(A)$ and $p_T(A)$ were not as strong ($r = .46, p < .0001$ for type b faces, $r = .51, p < .0001$ for type g faces). There was a very small negative correlation between the interference effects produced by the two different types of faces ($r = -.16, p = .04$). In addition, $p(A|B)$ differed between the two types of faces.

There were large individual differences in the interference effects. Across all participants, the standard deviation of the interference effect equaled .19 and .17 for the b and g types of faces, respectively. We computed the chi-square test statistic for the difference, $p(A) - p_T(A)$, between two sample proportions for each participant. This allows us to examine the size of the interference effect, regardless of its sign. If the null hypothesis of "no interference effect" is correct for each participant, then these chi-squares statistics should be distributed according to a central chi-squared distribution with degrees of freedom equal to one (assuming statistical independence of the observations). However, as can be seen in Table 2, the frequency of large chi-squares is higher than predicted, and so the goodness of fit test between the observed and predicted distributions rejects the null hypothesis for both types of faces ($\chi^2(4) = 15.83, p < .005$ for type b faces, $\chi^2(4) = 33.85, p < .0005$ for type g faces). These results indicate that even though the mean of interference effects across participants (which would cancel each other out if the effects were in different directions) is not different from zero for the type g faces, still the size of the observed interference effects among participants is larger than expected under the null hypothesis.

Table 2
Predicted and observed frequencies of chi-square values within each quantile bin.

Quantile	Predicted	Obs – type b	Obs – type g
.25	42.25	25	20
.50	42.25	45	46
.75	42.25	58	54
.90	25.35	20	16
1.0	16.9	21	33

3.3. Discussion

For Experiment 1, we changed some of the procedures used by Busemeyer et al. (2009). We reduced the number of training blocks, counterbalanced the association between face features and types, and compared blocking versus randomly mixing C-alone and D-alone trials during the transfer phase. The new procedures allowed us to test the robustness of the observed interference effects using a larger sample of participants. The basic findings were sufficiently robust to be replicated with these procedural variations.

Experiment 1 replicated the positive mean interference effect found with the type b faces; we also replicated the lack of observed mean interference effect for the type g faces. However, the interference effect observed for the type b faces was smaller in Experiment 1 as compared to the original study. It is possible that the reduced training used in Experiment 1 weakened the effect. Experiment 1 also replicated the finding that the probability of attacking given the face was categorized as bad increased for b faces as compared to g faces. Furthermore, we found a very small but significant negative correlation between the interference effects between g and b faces.

We also discovered that for the type g faces, although the mean interference effect was not statistically different from zero, the sizes (disregarding sign) of these interference effects were larger than predicted if one assumed that there were no systematic interference effects for each person.

Recall that the Markov model predicts that the probability of taking an action depends only on the category and not on the face. The results from Experiment 1 are clearly inconsistent with this property, as $p(A|B)$ differed between the two types of faces. Furthermore, the Markov model satisfies the law of total probability, which is inconsistent with the positive interference effect obtained with the type b faces. One could argue that the task of categorizing a face generates more attention to the categorization task and changes the probability of categorization between C-D and D-alone trials. However, this would produce an effect in the wrong direction – an increase of attention would increase the probability of a correct categorization for the C-D trials, which would increase the total probability to attack with respect to the D-alone trials.

The signal detection model also has difficulty accounting for the interference effects observed in Experiment 1. If the decision boundaries are the same for C-D and D-alone trials, then no interference is predicted, which is inconsistent with the results for the type b faces. If the bounds change so that a category response contracts the attack boundary, then we should have obtained a positive mean interference effect for both the type b and type g faces, but we did not observe it for type g faces. Furthermore, there should be a positive correlation across participants between the two interference effects, but the data showed a small negative effect instead.

Like the Markov model, the quantum model predicts that the probability of taking an action should depend only on the category and not the type of face, which is inconsistent with the results of Experiment 1. Unlike the Markov model, the quantum model can account for the interference effect, but the simple quantum model as described so far has difficulty accounting for the difference in interference effects for the two types of faces. When examining the means across participants, the interference appeared only with type b faces.

4. Experiment 2

In Experiment 1, on the C-D trials, the category response made by the person provided no new information regarding the probability of being rewarded for an action beyond what was already known from the face stimulus. For example, the probability of a reward for attacking given that a type b face was present was .54, and this did not change depending on whether the person categorized this face as good or bad, that is, $p(reward\,A|b, B) = p(reward\,A|b) = .54$. Therefore, the person did not learn anything new from his or her categorization about the best action to earn a reward. However, if in an experiment, the participant is actually told the category assignment of a face, then this would provide new information about the reward produced by each action. For example, the probability of a reward for attacking given the face is assigned to be a Lork (the bad guy) equals .70.

The empirical findings revealed in Experiment 1 suggest that participants treated their categorization response as new information – it was as if they were told the actual category assignment. To

test this idea, Experiment 2 introduced new transfer test trials, called X-D trials, in which the participant was not asked to make any categorization response and instead, the computer program identified the category of a face before requesting an action decision. This new type of transfer test trial provides a comparison of the action probabilities conditioned on the category between C-D trials and X-D trials. If participants treat their own categorization responses as if they were told the category assignment, then we should obtain no differences between the conditional probabilities from X-D versus C-D trials.

Also importantly, we can compare the interference effect produced by X-D and C-D types of trials, for which the Markov, signal detection, and quantum models generate different predictions. The X-D trials are similar to the D-alone trials with respect to the fact that participants were required only to make a single action decision in both cases. No categorization response was required from the participant on X-D trials. The interference effect for X-D trials is defined as the difference between the probability to attack on D-alone trials and the total probability to attack on X-D trials, with the latter estimated by disregarding the category assignment and computing the proportion of attack choices pooled across all X-D trials.

According to the Markov model, the probability of transiting to an attack action from a category state (e.g., $B \rightarrow A$) only depends on the category state (e.g., B). On C-D trials, the categorization response identifies the categorization state of a person; on X-D trials, the category assignment determines the categorization state of a person. Therefore, the probabilities of actions, conditioned on categories (e.g., $p(A|B)$ and $p(A|G)$) obtained on X-D trials should be equal to those obtained on C-D trials.

Although the Markov model does not predict interference effects for the comparison of C-D with D-alone trials, it does predict an interference effect for the comparison of X-D with D-alone trials. This is because the total probability to attack on the X-D trials is based on the probability that the experimenter assigns a face to a category, whereas the probability to attack on D-alone trials is presumably based on the participant's probability of categorizing a face.

According to a signal detection model, the probability of an attack action depends only on whether or not the face is located within the attack region, that is, $f \in R_A$, of the face space. Define C_B as the event that a face is assigned on X-D trials to the bad category, and define C_G as the event that a face is assigned on X-D trials to the good category. The probability of an attack decision conditioned on a bad category assignment equals $p(A|C_B) = p(f \in R_A|C_B)$, and likewise, the probability conditioned on the good category assignment equals $p(A|C_G) = p(f \in R_A|C_G)$. Together, these assumptions imply that $p(A) = p(f \in R_A) = p((C_G \cap (f \in R_A)) \cup (C_B \cap (f \in R_A))) = p(C_G) \cdot p(f \in R_A|C_G) + p(C_B) \cdot p(f \in R_A|C_B) = p_T(A)$. Therefore, the signal detection model predicts no interference when comparing D-alone to X-D trials.

Like the Markov model, the quantum model predicts that the probabilities of actions, conditioned on the category, obtained on X-D trials should be equal to those obtained on C-D trials. Also, for the same reason as the Markov model, the quantum model predicts interference effects for X-D trials. However, unlike the Markov model, the quantum model also predicts interference effects for C-D trials, as described earlier.

4.1. Method

4.1.1. Participants

The participants were 286 undergraduate students recruited from the same U.S. Midwest university for course extra credit. Of them, 59.44% were female; 80.07% were Caucasian, 7.34% were African American, 6.64% were Asian, and the rest identified themselves as "mixed" or "other." The average age was 20.46 (SD = 3.05).

4.1.2. Face stimuli

The face stimuli were the same as those used in Experiment 1 except for the following change. In Experiment 1, the association between face features and face types was a fixed percentage within each block. For example, of the 17 round face stimuli in each block, a randomly selected 60% were assigned to the Adok (the good guy) category, and of those Adok faces in each block, a randomly selected 70% were assigned to be friendly. Differently, Experiment 2 assigned the categorization and decision feedback probabilistically on each trial. For example, each round face stimulus had a .60 probability to be assigned to the Adok category, and each Adok face had a. 70 probability to be assigned to be friendly. Therefore, the correct category for a face stimulus could change across the blocks because of the probabilistic nature of the assignment.

4.1.3. Experimental procedure

The experimental procedure was the same as that of the second condition of Experiment 2 except for the following changes. A new type of transfer test trial, called the X-D trial, was added. On X-D trials, a face was shown, but no categorization response was requested. Instead, the computer disclosed the categorization assignment before the action decision question was prompted. Each participant completed four blocks of trials. Block 1 presented 34 C-D trials, and Block 2 comprised a mix of 34 X-D trials and 34 D-alone trials. After a 5-min break, Block 3 presented another 34 C-D trials, and Block 4 tested another mix of 34 X-D trials and 34 D-alone trials. As in previous studies, faces were randomized across trials within a block. In addition, to add time pressure, the time limit for answering categorization and decision questions was reduced from 10 s in Experiment 1 to 5 s in Experiment 2. For analysis, data from Blocks 3 and 4 were compared in the report below, treating Blocks 1 and 2 as training and practice trials. (Similar results are obtained if we compare Blocks 1 and 3 pooled together with Blocks 2 and 4 pooled together.)

4.2. Results

The estimated choice probabilities (i.e., sample proportions) were obtained for each participant and type of face from Blocks 3 and 4. Table 3 contains the results from Experiment 2. The rows labeled "Obs" present the observed findings, and the other rows present model predictions discussed later. Approximately 15% of the 286 participants were optimizers. Note that the probability to attack conditioned on the bad guy category, $p(A|B)$, increased on X-D trials compared to C-D trials. Also note that both C-D and X-D trials produced approximately the same positive interference effects for the type b faces, but they produced different

Table 3
Observed proportions and predicted probabilities from the Markov BA model and the quantum BAE model for Experiment 2.

	Cond	C-D			X-D		D	Int	Int
		$p(G)$	$p(A\|G)$	$p(A\|B)$	$p(A\|G)$	$p(A\|B)$	$p(A)$	C-D	X-D
Obs	70%, b	.24	.37	.61	.40	.69	.60	.04	.03
M	70%, b	.23	.39	.66	.39	.66	.59	.00	.05
Q	70%, b	.21	.33	.68	.41	.71	.63	.03	.04
Obs	70%, g	.78	.33	.53	.28	.58	.37	.00	−.03
M	70%, g	.77	.31	.56	.31	.56	.37	.00	−.04
Q	70%, g	.79	.33	.67	.32	.68	.40	.00	−.06

Note: Obs = observed, M = the Markov BA model, and Q = the quantum BAE model. The last two columns show the interference effects computed from C-D and X-D types of trials.

interference effects for the type *g* faces: There was no interference effect on the C-D trials but a negative interference effect on the X-D trials.

A *t*-test was performed separately for each face type, using the interference effect obtained from each participant as the dependent variable. When examining the mean interference effect across *all* participants (including optimizers), the effect for the type *b* faces was significant in the positive direction on both the C-D trials, $t(285) = 3.32, SE = .011, p = .001$, and the X-D trials, $t(285) = 4.33, SE = .010, p < .0005$. The effect for the type *g* faces was not significant on the C-D trials, but was significant in the negative direction on the X-D trials, $t(285) = -3.36, SE = .010, p = .001$. [3]

To summarize the interference effects from the C-D versus D-alone comparison, we computed summary statistics using all 455 participants from Experiments 1 and 2. The mean interference effect was .035 (SD = .19) for the type *b* faces, and it was .008 (SD = .17) for the type *g* faces. The 95% confidence interval of the mean interference effect ranged from .018 to .053 for the type *b* faces and ranged from −.008 to .023 for the type *g* faces. There was a moderately strong correlation between the estimates of $p(D|f)$ and $p_T(D|f)$, and it was higher for the *g* faces than for the *b* faces ($r = .52, p < .0001$ for the *b* faces; $r = .58, p < .0001$ for the *g* faces). The correlation of the interference effects between *b* and *g* faces was slightly negative ($r = -.10, p = .0275$).

4.3. Discussion

Experiment 2 introduced a new type of categorization-decision trial, the X-D trial. Unlike the C-D trials that asked the person to categorize each face, the X-D trials simply informed the participant about the category assignment of a face. The participant only had to make a decision, as on the D-alone trial. This provides comparisons of both C-D and X-D trials with D-alone trials when presented with type *g* and type *b* type of faces.

For the C-D trials, Experiment 2 replicated what was found in previous experiments: the positive interference effect obtained with the type *b* faces and the lack of interference effect for the type *g* faces. Also replicating Experiment 1, the probability to attack after categorizing the face as bad increased for the type *b* faces as compared to the type *g* faces (.61 vs. .53).

For the X-D trials, a new positive interference effect was obtained for the type *b* faces, but also a negative interference effect was found for the type *g* faces. Another interesting finding concerns the probability to attack conditioned on a bad guy categorization: This probability increased for the type *b* faces as compared to the type *g* faces (.69 vs. .58), as in the C-D trials. However, comparing the C-D and X-D trials, the probability to attack given that the face was categorized as bad was higher for X-D as compared to C-D trials. Note that the total probability to attack remained about the same for C-D and X-D trials.

Both the Markov and quantum models predict that the probability of an action, conditioned on the category, should be the same across both types of faces as well as across both C-D and X-D trials. However, the results indicate that the probability to attack conditioned on the bad guy category changed across face types (*b* vs. *g*) as well as trial types (C-D vs. X-D), ranging from .53 for type *g* faces during C-D trials to .69 for type *b* faces during X-D trials.

Both the Markov and quantum models predict interference effects for the X-D trials, which were observed. However, only the quantum model predicts interference effects for the C-D trials, which were observed for the type *b* faces (but not the type *g* faces).

The signal detection model implies no interference effects on X-D trials, but contrary to this prediction, positive interference occurred for type *b* faces and negative interference occurred for type *g* faces. To make the signal detection model account for the interference effects on X-D trials, we need to assume that the bound for attacking contracts for the bad category assignment and it expands for the good category assignment. This arbitrary change in bounds does not logically follow from signal detection theory because the category assignment provides restrictions on the sampling of faces in the multi-dimensional space – for example, the category assignment to a bad face should change the prior probability $p(f \in R_A)$ to a posterior probability $p(f \in R_A|C_B)$ rather than changing R_A itself. However, as we proved earlier, the latter assumption leads to the prediction of no interference.

5. Experiment 3

To further differentiate the competing models, in particular to differentiate them based on quantitative model comparisons, Experiment 3 included a new manipulation of the probability of the reward conditioned on the category. In Experiments 1 and 2 as well as earlier experiments by Townsend et al. (2000) and by Busemeyer et al. (2009), the probability of reward for attacking a face that was assigned to the bad guy category was .70, and likewise the probability of reward for withdrawing from a face that was assigned to the good guy category was also .70. Experiment 3 included two new conditions that varied the probability of reward: One condition used a lower .60 probability, and the other used a higher .80 probability. This manipulation was expected to change the certainty or uncertainty for action decisions, and change the action probabilities conditioned on the category. It provides an examination of the interference effects at different reward rates (i.e., uncertainty levels involved in the action decision).

5.1. Method

5.1.1. Participants

In total, 266 students from the same university participated in the experiments for course extra credit. They were similar to those in the preceding experiments. They were 20.12 years old on average (SD = 1.12), and 58.65% were female.

5.1.2. Face stimuli

The face stimuli were the same as those used in Experiment 2. The only difference from Experiment 2 was the reward rate for action decisions described earlier: Instead of the .70 probability of reward, .60 and .80 probabilities were used.

5.1.3. Experimental procedure

The experimental procedure was the same as that used in Experiment 2 except that the reward rate was changed. Participants were randomly assigned to one of the two reward probability conditions: 129 were assigned to the .60 condition, and 137 were assigned to the .80 condition.

5.2. Results

The estimated choice probabilities (i.e., sample proportions) were obtained for each participant and face type from Blocks 3 and 4, as the first two blocks were training and practice. Table 4 presents the results from all participants. (Only approximately 5% of the 266 participants were optimizers on the C-D trials for either

[3] Robert Nosofsky replicated Experiment 2 using 18 participants from Indiana University who received six blocks of training and two transfer blocks. For the type *b* faces, he found an interference effects equal to .11 and .04 for the C-D and X-D trials, respectively; for the type *g* faces, he found zero interference effects.

type of face, whose data could not be included in Table 4 because the marginal probability for the non-chosen category was zero, which prevents calculating the conditional probabilities and the total probability.)

The mean interference effects for the C-D versus D-alone comparison under the condition of the low reward rate of .60 equaled .03 for the type *b* faces and .00 for the type *g* faces, and both of these estimates lie within their respective confidence intervals estimated from the first two experiments; however, the interference effect from the high reward rate of .80 for the type *b* faces was close to zero, which is outside the confidence interval for this type of face from the first two experiments. The findings for the X-D condition for both reward rates are similar to those found in Experiment 2 – positive interference for the type *b* face and negative interference for the type *g* face.

A *t*-test was performed, separately for each face type, using the interference effect obtained from each participant as the dependent variable. The tests were computed using all participants. For the type *b* faces, the interference effect was significant in the positive direction under the .60 reward rate condition for both the C-D comparison ($t(128) = 2.14, SE = .014, p = .034$) and the X-D comparison ($t(128) = 5.11, SE = .011, p < .0005$); however, for the .80 reward rate, the effect was significant only for the X-D condition ($t(136) = 3.32, SE = .013, p = .001$). For the type *g* faces, the interference effect was significant in the negative direction only for the X-D condition under the .80 reward rate ($t(136) = 4.74, SE = .014, p < .0005$).

5.3. Discussion

Experiment 3 changed the reward rate for the appropriate action from .70 that was used in previous experiments to .60 for one group of participants and .80 for another. The .60 reward rate fairly closely replicated the results obtained earlier using the .70 reward rate. Increasing the reward rate to .80 generally increased the probability to attack for the bad guy category and decreased the probability to attack for the good guy category. In other words, participants' behavior became closer to optimal under the higher reward rate (i.e., with less uncertainty involved in action decisions). The increase in reward rate to .80 eliminated the positive interference effect for the type *b* faces on the C-D trials, but it also produced a negative interference effect for the type *g* faces during the X-D trials.

The results of Experiment 3 make it increasingly difficult to apply the signal detection model. If the bounds of the model remain unchanged across C-D, X-D, and D-alone trials, then no interference is predicted at all. If the bounds change for all of these trial types, then they must change in different directions for different trial types and for different reward rates. However, a rationale for all of these changes is lacking, making it difficult to formulate a coherent signal detection model that can be quantitatively fit to these data. Perhaps this is possible, but at present we do not have a clear way to build such a model. For example, although the stimuli varied according to face width and lip thickness, these two dimensions were perfectly correlated, and so the stimuli essentially varied according to one relevant dimension that we can interpret as face width. The distribution of face widths within each face type was unimodal and the two distributions were clearly separated, for example, with the good guy category faces positioned on one extreme end of the face widths continuum (say, e.g., the wide end of the face widths). Assume that there is a single cutoff on the face width dimension for categorizing good versus bad, and there is another single cutoff on this dimension for choosing withdrawing versus attacking. If the criterion for attacking on the face width dimension falls below the criterion for bad guy faces, then $p(A|G) = 0$ because these two events are mutually exclusive under these assumptions; for the same reason, if the criterion for bad guy faces falls below the criterion for attacking, then $p(W|B) = 0$. However, we observed zero choice probabilities from no individuals except for a small number of optimizers. Therefore, we would have to assume a higher dimensional space and a more complex set of boundaries, and then we would need to change them all in some *ad hoc* manner across X-D, C-D, and D-alone trial types and reward rates.

Table 4
Observed proportions and predicted probabilities from the Markov BA model and the quantum BAE model for Experiment 3.

	Cond	C-D			X-D		D	Int	Int
		$p(G)$	$p(A\|G)$	$p(A\|B)$	$p(A\|G)$	$p(A\|B)$	$p(A)$	C-D	X-D
Obs	60%, *b*	.24	.33	.66	.41	.67	.62	.03	.06
M	60%, *b*	.23	.47	.55	.47	.55	.53	.00	.01
Q	60%, *b*	.21	.32	.69	.43	.68	.63	.02	.05
Obs	60%, *g*	.77	.34	.58	.30	.57	.39	.00	−.02
M	60%, *g*	.77	.37	.46	.37	.46	.39	.00	−.02
Q	60%, *g*	.79	.32	.68	.34	.66	.40	.00	−.07
Obs	80%, *b*	.25	.26	.75	.30	.81	.64	.00	.04
M	80%, *b*	.23	.31	.77	.31	.77	.66	.00	.08
Q	80%, *b*	.21	.33	.68	.40	.74	.63	.02	.03
Obs	80%, *g*	.77	.23	.69	.19	.72	.33	.01	−.07
M	80%, *g*	.77	.26	.68	.26	.68	.36	.00	−.07
Q	80%, *g*	.79	.33	.67	.29	.71	.40	.00	−.06

Note: Obs = observed, M = the Markov BA model, and Q = the quantum BAE model. The last two columns show the interference effects computed from C-D and X-D types of trials.

Both the Markov and quantum models predict interference for the X-D condition, but only the quantum model can predict an interference effect for the C-D condition. It is difficult to know how this interference effect changes across the 60% and 80% reward rates without generating quantitative predictions based on parameter estimates. Although the Markov model fails to account for the positive interference effect on C-D trials, the effect is small and occurs only in one of the four conditions for C-D trials. It is quite possible that even though the quantum model can predict the interference for the one C-D condition, it may not be able at the same time to predict no interference in the other conditions, and so it remains unclear whether the Markov or the quantum model provides a better quantitative account of all the empirical results. Therefore, in the next section, we present a more rigorous quantitative comparison between generalized versions of the Markov and quantum models to determine which provides the better account of these new results.

6. Quantitative model comparisons

6.1. Model assumptions

As noted earlier, the Markov and quantum models are similar in many ways, but they also differ in some fundamental aspects. This section presents two generalized versions of the Markov and quantum models, side by side, in a parallel manner to clarify exactly where these two models differ. Both models are designed to describe how a person forms beliefs about a state of the world and decides to take actions under different states of the world. In other words, they both can be called belief-action (BA) models. A unique aspect of the quantum model is that it incorporates a concept of entanglement from quantum theory, and thus we call the quantum model a belief-action entanglement (BAE) model. The entanglement feature of the quantum model is discussed later. To begin, both models use a representation that has four *basis states* $\{GA, GW, BA, BW\}$, where, for example, *GW* symbolizes the

combined event of categorizing the face as a good guy and deciding to withdraw.

6.1.1. State representation

According to the Markov model, at each moment in time, the cognitive system is located in precisely one basis state (e.g., the basis state *GW*), but we (as theorists) cannot directly know another person's internal state, and hence assign a probability (e.g., ϕ_{GW}) that the cognitive system is in a state. Therefore, the Markov model assigns a probability to each basis state to produce a 4×1 probability distribution, $\phi = (\phi_{GA}, \phi_{GW}, \phi_{BA}, \phi_{BW})^T$, over the four basis states, which sums to one. This probability distribution is called a *mixture* state – the choice of a category and decision is determined by the person's current basis state; choice uncertainty arises from the theorist's lack of knowledge about a person's state.

According to the quantum model, at each moment in time, the cognitive system has a *potential*, called an *amplitude*, assigned to each basis state. The four belief and action potentials form a 4×1 *superposition* state $\psi = (\psi_{GA}, \psi_{GW}, \psi_{BA}, \psi_{BW})^T$. The superposition state has a very different conceptual meaning and interpretation in quantum theory: At each moment in time, the cognitive system is *not* exactly located in any one of the four basis states. In other words, the person is uncertain about his or her state and cannot say he or she is precisely in any single basis state. Only at the moment when a decision is required and the person selects a basis state does the person become located in this selected basis state.

Mathematically, the quantum superposition state ψ also operates differently from the Markov mixed state ϕ. For the Markov model, the current basis state is used by the person to determine an answer. Therefore, the probability that we as theorists assign to states, ϕ, directly determines the predicted probability of an answer. For the quantum model, the relation is less direct. Based on quantum theory, the answer that we observe from a person in a superposition state is indeterminate, and the probability of an answer equals the squared magnitude of an amplitude assigned to that basis state. For example, $|\psi_{GW}|^2$ represents the probability that the person categorizes the face as good and decides to withdraw. The state vector ψ is assumed to be unit length, $||\psi|| = 1$, so that the squared magnitudes of the four amplitudes sum to one.

6.1.2. Categorization process

Initially, after the presentation of a face but before any categorization or decision has been made, the information about the face is represented by an initial state. For the Markov model, the initial state is denoted as ϕ_f. For a type *b* face, we set $\phi_f = \phi_b = \frac{1}{2}(1 - p_C, 1 - p_C, p_C, p_C)^T$, where p_C is the probability that the person categorizes this type *b* face as a bad guy. Given this state, the probability of categorizing the face as a bad guy equals $\phi_{BA} + \phi_{BW} = p_C$. Similarly, for a type *g* face, we set $\phi_f = \phi_g = \frac{1}{2}(p_C, p_C, 1 - p_C, 1 - p_C)^T$, in which case the probability of categorizing the face as a good guy equals $\phi_{GA} + \phi_{GW} = p_C$. Note that according to the Markov model, after a face is presented, the person either thinks the face is good or thinks it is bad, but we as theorists do not know exactly which internal cognitive state of the person is present, and so we assign these initial probabilities to the cognitive states that the person may be in.

For the quantum model, the initial state is denoted ψ_f. For a type *b* face, we set $\psi_f = \psi_b = \frac{1}{\sqrt{2}}(\sqrt{1 - p_C}, \sqrt{1 - p_C}, \sqrt{p_C}, \sqrt{p_C})^T$, where p_C is the probability that the person categorizes this type *b* face as a bad guy. Given this state, the probability of categorizing the face as a bad guy equals $|\psi_{BA}|^2 + |\psi_{BW}|^2 = p_C$. Similarly, for a type *g* face, we set $\psi_f = \psi_g = \frac{1}{\sqrt{2}}(\sqrt{p_C}, \sqrt{p_C}, \sqrt{1 - p_C}, \sqrt{1 - p_C})^T$, in which case the probability of categorizing the face as a good guy equals $|\psi_{GA}|^2 + |\psi_{GW}|^2 = p_C$. Note that at this moment, the category of the face is unresolved, and so the person is not located in either category, but instead remains superposed between the two, and both have some potential to be expressed at that moment. On D-alone trials, when no categorization information or response occurs, the person remains in this initial state ψ_f, which is superposed with respect to the two categories. Only when we ask for a category on a C-D trial or when we tell the category on an X-D trial does the person resolve the uncertainty about the category. This idea based on quantum theory captures the fuzzy, uncertain feelings toward the face categories on the D-alone trials when there is no categorization response required or there is no categorization information provided.

On a C-D trial, the state is updated after the person categorizes the face; likewise on an X-D trial, the state is updated after the person is told the category.[4] If the person categorizes the face as a bad guy, then the state is updated to be consistent with this categorization response. For the Markov model, the mixed state is updated to $\phi_f \rightarrow \phi_B = \frac{1}{2}(0, 0, 1, 1)^T$ because we have recorded the person's current beliefs and we now know which category the person is thinking about. For the quantum model, the superposition state is updated to $\psi_f \rightarrow \psi_B = \frac{1}{\sqrt{2}}(0, 0, 1, 1)^T$, so that the probability of categorizing the face as a bad guy becomes 1.0 afterwards. Likewise, if the person categorizes the face as a good guy, then it is updated to be consistent with this categorization response. For the Markov model, it is updated to $\phi_f \rightarrow \phi_B = \frac{1}{2}(1, 1, 0, 0)^T$; for the quantum model, it is updated to $\psi_f \rightarrow \psi_G = \frac{1}{\sqrt{2}}(1, 1, 0, 0)^T$. The updates for the X-D trials are computed in the exact manner as those for the C-D trials.

6.1.3. Action evaluation process

At this point, if a categorization was made first, then the person is in one of the states, ϕ_B or ϕ_G, for the Markov model; the person is in one of the states, ψ_B or ψ_G, for the quantum model. If no categorization was made first, then the person remains in the initial state: ϕ_f for the Markov model and ψ_f for the quantum model. However, no evaluation of actions has yet occurred, and the two actions are initially equally likely. During the action evaluation stage, the previously unbiased state is transformed to favor one or the other action. This transformation depends on the utilities of the payoffs for each category and action. If the bad guy category is believed to be present, then the state is transformed to some extent toward the attack action; if the good guy category is believed to be present, then the state is transformed to some extent toward the withdraw action.

Technically, the transformation for the Markov model is computed by using a 4×4 transition matrix *T*. The element T_{ij} in row *i* and column *j* of *T* represents the probability of transiting to basis state *i* from basis state *j*. The transformation for the quantum model is computed by using a 4×4 unitary matrix, denoted *U*. The element U_{ij} in row *i* and column *j* of *U* represents the amplitude for the transition to the basis state *i* from the basis state *j*; the probability of this transition equals the squared magnitude of the amplitude. (See Appendix B for mathematical details concerning the construction of these matrices from utility parameters described below.)

It is again helpful to compare the interpretations of the Markov versus quantum model. According to the Markov model, the person's cognitive system moves from exactly one basis state to another to produce a trajectory of basis states across time, and the Markov process describes the probability that a person follows

[4] This is related to the quantum "collapse" of the wave function that follows a measurement.

a particular trajectory. According to the quantum model, there is not any single, particular trajectory of basis states across time, and instead the superposed basis states themselves evolve across time until a decision is made, upon which the person becomes located in a specific basis state created by the decision.

6.1.4. Action selection process

Finally, for the Markov model, if ϕ is the mixed state before evaluating the actions, then $\phi_F = T \cdot \phi = (\phi_{GA}, \phi_{GW}, \phi_{BA}, \phi_{BW})^T$ is the transformed state after evaluating the actions. According to the Markov model, the person is always located in some particular basis state; if the person is located in either state *GA* or *BA* immediately *before* the time of decision, then at the time of decision the person chooses to attack. Again, we (as theorists) are uncertain about the state the person is located in, but we can assign probabilities, ϕ_F, that the person's cognitive system is in each of the four basis states at the time of decision. For example, our prediction for the probability that a person decides to attack equals $\phi_{GA} + \phi_{BA}$.

For the quantum model, if ψ is the state before evaluating the actions, then $\psi_F = U \cdot \psi = (\psi_{GA}, \psi_{GW}, \psi_{BA}, \psi_{BW})^T$ is the transformed state after evaluating the actions. According to the quantum model, immediately before the time of decision, the person is in a superposition state. At the time of decision, the person must resolve this indeterminacy. If the person becomes resolved on either the *GA* or the *BA* state, then the person chooses to attack. Therefore, using the final evaluation state, ψ_F, the probability of deciding to attack equals $|\psi_{GA}| + |\psi_{BA}|^2$.

In sum, if the face was first categorized as a good guy, then the evaluation state $\phi_F = T \cdot \phi_G$ is used for the Markov model and $\psi_F = U \cdot \psi_G$ is used for the quantum model to compute the action probabilities; if the state was first categorized as a bad guy, then the evaluation state $\phi_F = T \cdot \phi_B$ is used for the Markov model and $\psi_F = U \cdot \psi_B$ is used for the quantum model to compute the action probabilities; and if no categorization was made, then $\phi_F = T \cdot \phi_f$ is used for the Markov model and $\psi_F = U \cdot \psi_f$ is used for the quantum model to compute the action probabilities.

6.2. Model parameters

6.2.1. Initial state

For both models, the initial state contains one parameter, p_C, representing the probability of "correctly" (i.e., optimally) categorizing a face, that is, categorizing a type *b* face as a bad guy and categorizing a type *g* face as a good guy. We assume that this parameter is the same across both types of faces (and this assumption is consistent with the average results in our experiments).

6.2.2. Utilities for actions

Recall that the transition and unitary matrices of the Markov and quantum models, respectively, represent the evaluation of payoffs for determining the probability of taking each action. First of all, these evaluations depend on the category, because participants are more frequently rewarded for attacking faces categorized as bad and they are more frequently rewarded for withdrawing from faces categorized as good. Second, these evaluations depend on the rate of reward, denoted here as *R*, which was 70% in Experiments 1 and 2, and it varied between 60% and 80% in Experiment 3. Finally, we also need to assume that the utilities of actions change for the different types of faces. As suggested by behavioral research on prejudice and stereotyping (e.g., Allport, 1954; Devine, 1989; Dovidio, Hewstone, Glick, & Esses, 2010), humans' reactions to others are affected by stereotypes and bias. The association between face features (e.g., face shapes) and face types (*b* vs. *g*) manipulated in the experiments can be viewed as stereotypes and bias. Participants may feel "right" and more justified (i.e., positive utilities) when attacking a bad guy type of face as compared to a good guy type of face, and they may feel it to be "wrong" and less justified to withdraw from a bad guy type of face. Likewise, they may feel "wrong" and less justified (i.e., negative utilities) when attacking a good guy type of face as compared to a bad guy type of face, and they may feel it is fairer to withdraw from a good guy type of face. That is, the utilities of actions differ between the two types of faces (*b* vs. *g*).

6.2.3. Markov transition matrix

The Markov BA model uses a separate 4×4 transition matrix T for each type of face (*b* vs. *g*). For a given type of face, each transition matrix has two transition rates (α_G, β_G) that apply for the good guy categorization, and another two transition rates (α_B, β_B) that apply for the bad guy categorization. The transition rate α determines transitions from attack states to withdraw states, and β determines transitions from withdraw states to attack states. The probability to choose attack is an increasing function of the ratio, $k = \beta/\alpha$.

The reward rate R is the probability to reward withdraw actions towards good guys. For good guys, it is included by multiplying α by the probability of being rewarded for withdrawing (R) and by multiplying β by the probability of being rewarded for attacking ($1-R$). For the good guy category, this produces transition rates $(R \cdot \alpha_G, (1-R) \cdot \beta_G)$, so that the probability to attack is an increasing function of $\left(\frac{1-R}{R}\right) \cdot \left(\frac{\beta_G}{\alpha_G}\right) = \left(\frac{1-R}{R}\right) \cdot k_G$. Following the same idea, for the bad guy category, including the reward produces transition rates $((1-R) \cdot \alpha_B, R \cdot \beta_B)$, and in this case the probability to attack is an increasing function of $\left(\frac{R}{1-R}\right) \cdot \left(\frac{\beta_B}{\alpha_B}\right) = \frac{R}{(1-R)} \cdot k_B$. Therefore, two rate parameters – k_G for the good guy category and k_B for the bad guy category – need to be estimated for each type of face.

In sum, the Markov BA model for both types of faces requires fitting 4 parameters to the two transition matrices: (k_G, k_B) for the transition matrix applied to the type *b* face, and (k_G, k_B) for the transition matrix applied to the type *g* face. All of these 4 parameters are necessary for the Markov model. Imposing constraints would force the model to fail to predict important qualitative aspects of the data. Including the parameter p_C, there are 5 parameters in total. Appendix B describes the details for constructing the transition matrices from these parameters.

6.2.4. Quantum unitary matrix

The quantum BAE model uses a separate 4×4 unitary matrix for each type of face (*b* vs. *g*). For a given type of face, each unitary matrix has one utility μ_G representing the (negative) utility for attacking a face in the good guy category, and another utility μ_B representing the (positive) utility for attacking a face in the bad guy category. When the good guy category applies, the probability to attack is an increasing function of μ_G; when the bad guy categories applies, the probability to attack is an increasing function of μ_B. The reward rate is included by multiplying each utility parameter (μ_G, μ_B) by the reward rate R to produce $R \cdot \mu_G, R \cdot \mu_B$.

The quantum BAE model also introduces an important concept in quantum theory, entanglement. The psychological function of entanglement is to coordinate beliefs and actions. As suggested by research on cognitive dissonance (e.g., Festinger, 1957) and social projection (e.g., Busemeyer & Pothos, 2012; Krueger, DiDonato, & Freestone, 2012), participants feel the need to be consistent with their beliefs and actions. Entangling of beliefs with actions can occur on either D-alone trials or C-D trials, because the category remains unknown during the decision and beliefs about the category can change. Entanglement cannot occur on X-D trials, because the category is exposed before the decision,

causing beliefs to remain fixed on the known category during the decision.

An additional parameter γ is used in the unitary matrix to coordinate beliefs with actions to form the *entangled state*. Formally, an entangled state is a superposition state in which the amplitude assigned to each basis state for each combination of answers (e.g., *BA*) cannot be decomposed into a product of individual amplitudes for each answer (e.g., $\psi_{BA} \neq \psi_B \cdot \psi_A$). In the model, the entanglement process produces a state in which amplitudes associated with consistent beliefs and actions, ψ_{GW}, ψ_{BA}, are enhanced, and amplitudes associated with inconsistent beliefs and actions, ψ_{GA}, ψ_{BW}, are attenuated. The entanglement parameter is critical for producing interference effects: If $\gamma = 0$, then the BAE model does not produce any interference effects; only when γ is not zero, can the BAE model produce interference effects.

To be fair, we could include this entanglement concept in the Markov model. In fact, Busemeyer and Bruza (2012, p. 275), did implement this idea in the Markov model. However, even when the entanglement concept is included, the Markov model still predicts no interference (see Busemeyer and Bruza, p. 276). Moreover, including a non-zero entanglement parameter only decreases the fit of the Markov model.

Through an analysis of the BAE model, we found that the interference pattern differs for *g* versus *b* type faces because of the following interaction between the utility and entanglement parameters. The size of the interference produced by γ depends on the degree of asymmetry in the utility parameters. In particular, if the utilities are symmetric, that is, $\mu_G = -\mu_B$, then the interference disappears. Interference requires asymmetric utilities, $\mu_G \neq -\mu_B$. If $\gamma > 0$ and $\mu_B > -\mu_G$, then the interference tends to be positive. If $\gamma > 0$ and $-\mu_G > \mu_B$, then the interference tends to be negative. Changing the sign of γ reverses these relations. In other words, the BAE model generates positive interference when the utilities are asymmetric and $\mu_B + \mu_G$ has the same sign as the entanglement γ, and it produces negative interference when they have opposite signs. According to the BAE model, different interference effects for each face type are produced by the interaction between the entanglement parameter and the utility parameters. We constrained the quantum model to predict the basic pattern of interference effects by requiring $\gamma > 0$ and $\mu_B > -\mu_G$ for the type *b* faces and $\mu_B = -\mu_G$ for the type *g* faces. Accordingly, our primary hypothesis based on the quantum BAE model is that the change in the interference effects across face types is attributed to changes in the utilities for attacking and withdrawing from each type of face, while the entanglement parameter remains the same across types of faces.

In sum, the quantum BAE model for both types of faces entails 4 parameters for the two unitary matrices: one entanglement parameter γ, one utility parameter $(\mu_G, \mu_B = -\mu_G)$ for the type *g* faces, and two utility parameters (μ_G, μ_B) for the type *b* faces. Including the parameter p_C, there are 5 parameters altogether. Appendix B describes the details for constructing the unitary matrices from these parameters.

6.3. Model predictions

6.3.1. Model predictions for Experiment 1

The Busemeyer et al. (2009) experiment as well as Experiment 1 provide only 8 data points: 4 per face type, $[p(G), p(A|G), p(A|B), p(A)]$; note that $p(B) = 1 - p(G)$ and $p_T(A)$ are derived from the other probabilities. However, they still provide a challenge for the competing models. Using 5 parameters, the Markov model can accurately predict the 6 data points from the C-D trials; nevertheless, it predicts no interference effect for the type *b* face, which was observed in these experiments. Using 5 parameters, the quantum model also can produce accurate predictions for these experiments. The last two rows of Table 1 present an example of the predictions computed from the quantum BAE model using the constraints from our primary hypothesis about the model parameters. In particular, we set $p_C = .80$, $\gamma = 0.9120$ for both types of faces; we used $\mu_G = -0.0864$, $\mu_B = 0.4205$ for the type *b* face, and we used $\mu_G = -0.26, \mu_B = 0.26$ for the type *g* face. For both types of faces, the utility parameter for the good guy category is negative (i.e., lowering the probability of attacking), and the utility parameter for the bad guy category is positive (i.e., raising the probability of attacking). The utility parameters are asymmetric for the *b* type but symmetric for the *g* type. As can be seen in Table 1, the quantum model reproduces the pattern of the observed average findings for both types of faces in the experiments.

6.3.2. Model predictions for Experiment 2

Experiments 2 and 3 provide larger experimental designs and more data points to quantitatively compare the Markov and quantum models. Here using these experiments, we present the first strong quantitative test of these models by using a generalization criterion method (Busemeyer & Wang, 2000). A main advantage of the generalization criterion method of model comparision is its reliance on accurate *a priori* predictions of the models to new conditions, or theoretical extrapoliations of the models to new conditions (Busemeyer & Wang, 2000). More specifically in our case, first we fit the parameters to the 12 data points from Experiment 2 (reward rate *R* = 70%), and then we used these same exact parameters in a generalization test to predict the 24 data points for two new experimental conditions in Experiment 3 (reward rate *R* = 60%, 80%).[5]

First, we evaluated the Markov and quantum models' fits to the results from Experiment 2. The first six columns of Table 3 contain 3 independent data points $[p(G), p(A|G), p(A|B)]$ from the C-D trials, 2 independent data points $[p(A|G), p(A|B)]$ from the X-D trials, and 1 independent data point $[p(A)]$ from the D-alone trials, and so there are $2 \times 6 = 12$ data points to fit for both types of faces. We evaluated (1) the overall badness of fit of each model using a root mean squared error criterion, $RMSE = \sqrt{SSE/N}$, where *SSE* equals the sum of squared errors for each row and summed across rows, and *N* refers to the number of data points; and (2) goodness of fit using $R^2 = 1 - SSE/TSS$, where *TSS* = the total sum of squared deviations around the overall mean proportion.

For the Markov BA model, a total of 5 parameters were fit to the 12 data points in Table 3 using a least squares criterion (see Appendix B for the parameter values). The minimum *SSE* fit index produced $RMSE = .024$ and $R^2 = .99$. The overall fit to most of the data points is good, which is not surprising given the large number of parameters. Nevertheless, the Markov model cannot produce the positive interference effect for the type *b* face during the C-D trials (see Table 3).

For the quantum BAE model, a total of 5 parameters were fit to the 12 data points in Table 3 using a least squares criterion (see Appendix B for the parameter values). The minimum fit index produced $RMSE = .061$ and $R^2 = .97$. The overall fit to the data points is not as good as that of the Markov model. However, the quantum BAE model has the benefit of accounting for the positive interference effect for the type *b* faces and lack of interference for the type *g* faces during the C-D trials (see Table 3).

[5] If we freely fit all 5 parameters to all 36 data points from all three reward conditions (60%, 70%, 80%) from Experiments 2 and 3, then both models fit approximately equally well, producing $R^2 = .98$. The Markov model fits the choices conditioned on each category better, while the quantum model fits the interference effects better.

In summary, both models were fit to the 12 data points using the same number of parameters. Both models fit reasonably well, but the Markov model produced an overall better fit than the quantum model. Both models can account for the difference in probability to attack conditioned on the categorization between the type *b* and the *g* faces. Both models can account for the interference effects obtained on X-D trials. Only the quantum BAE model accounts for the positive interference effect for the type *b* faces (and also the absence of interference effect for the type *g* faces) during C-D trials.

6.3.3. Model predictions for Experiment 3

The data in Table 4 from Experiment 3 contains 24 free data points to test the predictions of the models: 12 contributed by the C-D trials, 8 contributed by the X-D trials, and 4 contributed by the D-alone trials. For both models, we used exactly the same parameters estimated from Experiment 2 (i.e., from the reward rate $R = .70$ condition) to make *a priori* predictions for the two new reward rate conditions (i.e., $R = .60$ and $R = .80$). Table 4 shows the results of the predictions. The Markov model produced fairly accurate predictions, with RMSE = .068 and $R^2 = .97$. The quantum model now produced more accurate predictions, with RMSE =.058 and $R^2 = .98$. Both models managed to correctly predict the pattern of interference effects produced by the X-D trials; however, only the quantum model predicted the correct pattern of interference effects for the C-D trials (with one exception being the interference effect for the $R = .80$ condition).

Based on the model comparisons, we make the following conclusions. The two models can fit a large data set of categorization-decision results reasonably well. The quantum BAE model has one unique advantage – that is, accounting for the pattern of interference effects obtained on C-D trials. This advantage of the quantum BAE model over the Markov BA model does not entail any disadvantage in terms of overall accuracy of model predictions.

7. General discussion

7.1. Summary of main empirical findings

This article investigated the relation between categorization and decision making using a new experimental paradigm. Participants were first shown a face. Under a C-D condition, they were asked to categorize it as either a good guy or bad guy and then decide to attack or withdraw; under an X-D condition, they were informed about the category first and then decided to attack or withdraw; under a D-alone condition, they simply decided to attack or withdraw. This paradigm allows an investigation of a phenomenon that we call an interference effect based on quantum theory, which is a type of violation of the important classical law of total probability. An interference effect is defined as the difference between (1) the total probability of deciding to take an action pooled across categories for C-D or X-D trials, as compared to (2) the probability of deciding to take the same action on D-alone trials.

Previously, Busemeyer et al. (2009) revealed a surprising finding regarding the interference of categorization on decision making. When comparing C-D and D-alone trials, a positive interference effect occurred with a type of face that was most frequently associated with the bad guy category (the type *b* faces), but no interference occurred with a type of face associated with the good guy category (the type *g* faces).

This article reports three new experiments that used (1) much larger samples of participants, (2) new variations in procedures, and (3) new experimental conditions to further explore this phenomenon. First, we found that the original interference effect obtained by comparing C-D with D-alone trials is robust across wide variations in procedures and across a large sample of participants (although we also found large individual differences in the effect). Second, we discovered a new type of interference effect produced by comparing X-D and D-alone trials: A positive interference effect occurred with the type *b* faces, and a negative interference effect occurred with the type *g* faces. Third, we discovered that increasing the probability of rewarding the appropriate action for a category decreases the positive interference effect obtained with C-D trials, but at the same time it increases the negative interference effects obtained on X-D trials.[6]

7.2. Theoretical implications

The observed interference effects interacted with the types of faces (*b* vs. *g*) and with the types of categorization-decision trials (C-D vs. X-D). This pattern of results presents challenges to traditional cognitive models. The multi-dimensional signal detection model (e.g., Ashby & Townsend, 1986) does not predict any interference effects at all for either C-D or X-D types of trials. Of course, it may be possible to come up with alternative explanations after these interference effects are known. In particular, to account for the interference effects using the signal detection model, one could make *post hoc* assumptions regarding changes in the decision boundaries. However, these assumptions would have to differ across types of faces and across types of categorization-decision trials, which makes this theory intractable for developing a cogent model to fit to these results. A Markov model, which was originally proposed for this categorization-decision paradigm (Townsend et al., 2000), cannot explain the pattern of the observed interference effects either. Although the Markov model can account for interference effects obtained on the X-D trials, it does not predict any interference effects for the C-D trials. Recently, Busemeyer et al. (2009) pointed out that a quantum model predicted *a priori* that interference effects could occur using the categorization-decision paradigm.

One could argue that although a quantum model can predict interference effects, doing so may come at a cost of making inaccurate predictions for other aspects of the data. Although the Markov model cannot predict interference effects, it may provide more accurate predictions for other aspects of the data. Therefore, for the first time within the categorization-decision paradigm, we conducted a rigorous quantitative comparison of the accuracy of Markov versus quantum models for fitting large experimental designs involving factorial manipulations of reward rates, face features and face types, and types of categorization-decision trials (Experiments 2 and 3) based on a generalization criterion method (Busemeyer & Wang, 2000). We fit the parameters to the data from Experiment 2 and then used these same exact parameters to *a priori* predict data from new experimental conditions in Experiment 3. The results of the model comparison demonstrated that the advantage of the quantum model to account for interference effects is not offset by a loss in accuracy for predicting the other aspects of the data as compared to the Markov model. Therefore, the quantum model not only accounts for the interference effects that cannot be predicted by the Markov model, but also fits other aspects of the data as well as the Markov model does. The quantum model predicts better than the Markov model in the generalization test.

The quantum BAE model employs the quantum principles of superposition and entanglement for explaining the psychological mechanisms underlying the puzzling violation of the classical

[6] Robert Nosofsky (2013) independently replicated these results as well; see Footnote 3.

law of total probability produced by interference effects. Specifically, the BAE model includes four psychologically meaningful parameters: (1) p_c, representing the probability of categorizing a type b face as bad (or a type g face as good); (2) one utility parameter $(\mu_G, \mu_B = -\mu_G)$ for the type g faces and two seperate utility parameters (μ_G, μ_B) for the type b faces, where for a given type of face stimulus, μ_G represents the (negative) utility of attacking a face that is categorized as good and μ_B represents the (positive) utility of attacking a face that is categorized as bad; and (3) γ, representing an entanglement parameter that coordinates beliefs with actions in a consistent manner. The parameters of the quantum model help provide a psychologically meaningful interpretation for why the interference effect occurs most strongly for the bad guy type of face and virtually disappears for the good guy type of face under certain conditions. Our primary hypothesis was that the change in the interference effects across face types is attributable to changes in the utilities for attacking and withdrawing from each type of face, even though the belief-action entanglement parameter remains the same across both types of faces. Our modeling results support this hypothesis.

7.3. Research on related paradigms

The categorization-decision task paradigm requires inferring a category and then making an action decision. An important feature of the categorization-decision task is the inclusion of consequences that depend on both the correct category and action being selected during the decision. Although the categorization-decision making paradigm is a new research topic, there have been three closely related experimental paradigms that have received investigation: (1) a categorization-categorization paradigm examining the effect of an initial categorization on a subsequent categorization, (2) a category-feature inference paradigm investigating inferences about features of an object following categorization of the object, and (3) a feature-feature inference paradigm investigating feature inferences based on reasoning from causal networks. These three paradigms, as described in detail below, are different from the categorization-decision paradigm investigated in the current article because they do not include a categorization stage that has effects on subsequent decision making.

Di Nunzio, Bruza, and Sitbon (2014) investigated a categorization-categorization task. Participants were asked to categorize documents, which were news articles that appeared in the Reuters newswire in 1987. The participants were recruited by the Mechanical Turk system, and they differed according to their expertise and experience with online experiments (masters vs. non-masters). A total of 82 documents were selected for the task. Each document could be categorized in two different ways: as being about "crude oil" or about "shipping." One group was given a two-step $C_O - C_S$ categorization task, and these participants first categorized the document with respect to "crude oil" and then categorized it with respect to "shipping." The other group was given a single step C_S-alone task, and these participants simply categorized the document with respect to "shipping." The results of this study revealed large positive interference effects, similar to our findings with the type b faces, for participants who were not "masters" of Mechanical Turk; but surprisingly, a negative interference effect occurred with the "master" level of expertise on Mechanical Turk.

Murphy and Ross (1994) and later Griffiths et al. (2012) investigated a categorization-inference task. Participants were shown several categories of objects; each category contained several different objects, and each object was described by three different binary valued features. One object was selected out of all the categories, and the participant was informed about one of the features of the selected object. Given this feature information, the participant was asked to infer the value of another unknown feature of the selected object. The general finding was that participants tended to first infer the most likely category of the selected object based on the known feature, and then infer the unknown feature solely on the basis of the previously inferred category. This finding violates the law of total probability, because participants should infer a feature according to a weighted average across all categories rather than focusing solely on the most likely category.

Chaigneau et al. (2004) as well as Rehder and Burnette (2005) investigated feature inferences based on descriptions of causal networks. Participants were provided with a description of a causal network (e.g., a causal chain $X \rightarrow Y \rightarrow Z$). Then they were asked to predict the occurrence of a feature (e.g., Z is present) under two conditions: (1) Make the inference conditioned on the presence of both a direct parent cause and another factor that affects the outcome by an indirect path only (e.g., infer Z given both Y and X); (2) make the inference conditioned only on the direct parent cause (e.g., infer Z given Y). If the inferences are identical across the two conditions, then "screening off" is said to be satisfied. However, participants often violated "screening off" because they continued to be influenced by the indirect factors. Violations of "screening off" also occurred in our categorization-decision paradigm: The type of face continued to influence the probability of taking an action even after the category of the face was selected by or revealed to the decision maker.

Apparently there are close connections between these three different lines of research and the current categorization-decision paradigm. Perhaps a common theoretical explanation could underlie them all. Recently, Nosofksy (2015) developed an exemplar model to account for the category – feature inference task. Future research is needed to examine the applications of Markov and quantum models to these related experimental paradigms, and to compare the accuracy of predictions of Markov and quantum models to that of the exemplar model recently proposed by Nosofksy (2015).

7.4. Broader perspectives

It is worth pointing out again that the quantum BAE model was originally developed for a completely different type of decision making task – a prisoner's dilemma game (Pothos & Busemeyer, 2009). In the prisoner's dilemma paradigm, a violation of the law of total probability occurred when comparing two conditions: (1) when the move of an opponent player was known ahead of time, versus (2) when the opponent's move was unknown. The psychological intuition used to explain the violations of "rational" decision making in the prisoner's dilemma paradigm is exactly like that in the current categorization-decision paradigm (Busemeyer & Pothos, 2012). Indeed, a strength of the BAE model is that it is based on a small set of coherent quantum probability rules and concepts instead of *ad hoc* assumptions, and it provides a unifying theoretical principle for explaining different psychological phenomena. In fact, the same set of quantum probability rules and concepts have been used to account for a large variety of puzzling findings, in domains ranging from judgment and decision (Busemeyer, Wang, & Shiffrin, 2015c; Busemeyer, Wang, & Townsend, 2006; Khrennikov & Haven, 2009; Yukalov & Sornette, 2011) to language and thinking (Aerts, Gabora, & Sozzo, 2013; Blutner, Pothos, & Bruza, 2013), and from casual reasoning (Trueblood & Busemeyer, 2012) to perception and memory (Atmanspacher & Filk, 2010; Brainerd, Wang, & Reyna, 2013; Brainerd, Wang, Reyna, & Nakamura, 2015). For example, similar models have been used to address "irrational" probability judgment errors, such as disjunction and conjunction fallacies (Busemeyer, Pothos, Franco, & Trueblood, 2011; Busemeyer, Wang, Pothos, & Trueblood, 2015b) and asymmetric similarity

judgments (Pothos, Busemeyer, & Trueblood, 2013; Pothos & Trueblood, 2014), attitude judgments (White, Pothos, & Busemeyer, 2014), as well as order effects on sequential judgments and decisions (Wang & Busemeyer, 2013, Wang & Busemeyer, 2016; Wang, Solloway, Shiffrin, & Busemeyer, 2014) and inferences (Trueblood & Busemeyer, 2011). In particular, the superposition concept has been a basic idea in many of these quantum cognition applications, for instance, providing a simple explanation for episodic memory over-distribution phenomena (Brainerd et al., 2013) and the interference of choice on later confidence (Kvam, Pleskac, Yu, & Busemeyer, 2015). So has the entanglement concept. The quantum entanglement-like behavior has been observed in human semantic networks, mental lexicons, and word associations (Aerts et al., 2013; Bruza, Kitto, Nelson, & McEvoy, 2009; Bruza, Kitto, Ramm, & Sitbon, 2015a, Bruza, Wang, & Busemeyer, 2015b) as well as ambiguous perception (Atmanspacher & Filk, 2010).

Quantum theory is still unfamiliar to most psychologists, but the emergence of the new field of quantum cognition is a call to address accumulating findings in cognition that have resisted coherent, principled classical explanations for decades (for a review, see Busemeyer & Bruza, 2012; Busemeyer, Wang, & Pothos, 2015a; Khrennikov, 2010; for a brief introduction, see Bruza et al., 2015a, 2015b; Busemeyer & Wang, 2015; Busemeyer, Wang, Khrennikov, & Basieva, 2014; Pothos & Busemeyer, 2013; Wang & Busemeyer, 2015; Wang, Busemeyer, Atmanspacher, & Pothos, 2013). In fact, some of the founding fathers of quantum theory, most ardently Niels Bohr, argued a century ago that quantum theory would prove useful for psychology and philosophy (Murdoch, 1987; Pais, 1991). It is an interesting twist of history that some of the key conceptions of quantum theory actually were proposed by psychologists before they proved essential for quantum physics (Murdoch, 1987; Pais, 1991), but it is quantum physics that rigorously formalized these concepts and developed a coherent mathematical foundation for the theory, which enable precise empirical predictions and testing. The new field of quantum cognition takes advantage of these abstract, mathematical principles of quantum theory (i.e., quantum probability theory) to formalize psychological states and process, which has proven fruitful for addressing many enduring psychological questions. In this article, we have shown how the BAE model utilizes the basic quantum concepts of superposition and entanglement, which have been used to account for the prisoner's dilemma, to formalize the psychological process leading to the interference effect of categorization on subsequent decision making. The quantum BAE model should be useful for examining a large range of applications. It can be applied to many categorization and decision tasks (e.g., medical diagnosis and treatment decisions, problem categorization and solution) and situations in which beliefs and actions are interdependent (e.g., White et al., 2014).

Acknowledgements

The work was supported by the US Air Force Office of Scientific Research (FA 9550-12-1-0397 and FA 9550-15-1-0343) and the US National Science Foundation (SES-1153846, SES-1153726) to both authors. We thank Tyler Solloway and Cody Cooper for assistance with some of the data collection, and we thank Editor-in-Chief Steven Sloman and three anonymous reviewers for their helpful comments.

Appendix A

Below, we prove that if the parameters remain the same across C-D trials and D-alone trials, then a very general class of Markov models must predict no interference effects (see also Busemeyer & Bruza, 2012, chap. 8). Suppose N is the dimension of the state space. For convenience, define $L^T = [1 \quad 1 \quad \cdots \quad 1]$ as a $1 \times N$ row vector containing all ones, which is used for summation. Define M_G as a $N \times N$ diagonal matrix with ones on the diagonal corresponding to states identified with the good guy category, and zeros otherwise; define M_B as a $N \times N$ diagonal matrix with ones on the diagonal corresponding to states identified with the bad guy category, and zeros otherwise. These two events are mutually exclusive and exhaustive, so that we require $M_G \cdot M_B = \mathbf{0}$ and $M_G + M_B = \mathbf{I}$. Define M_A as a $N \times N$ diagonal matrix with ones on the diagonal corresponding to states identified with the attack decision, and zeros otherwise; define M_W as a $N \times N$ diagonal matrix with ones on the diagonal corresponding to states identified with the withdraw decision, and zeros otherwise. These two events are mutually exclusive and exhaustive, so that we require $M_A \cdot M_W = \mathbf{0}$ and $M_A + M_W = \mathbf{I}$.

The presentation of a face produces an initial state, which in general is a $N \times 1$ column vector ϕ_I. On C-D trials, a categorization response is based on the initial state ϕ_I. The probability of categorizing as a good guy equals $p(G) = L^T M_G \phi_I$, and the probability of categorizing as a bad guy equals $p(B) = L^T M_B \phi_I$. The revised state, conditioned on the categorization as a good guy, equals $\phi_G = M_G \phi_I / p(G)$; the revised state, conditioned on the categorization as a bad guy equals $\phi_B = M_B \phi_I / p(B)$. On D-alone trials, no categorization occurs, and so the state remains at the initial state ϕ_I; however, it is useful to rewrite this initial state as a mixture of the two conditional states: $\phi_I = p(G)\phi_G + p(B)\phi_B$.

Now suppose that the actions are evaluated on the basis of the payoffs, and the states are transformed by a general $N \times N$ transition matrix T. On a C-D trial, following a good guy categorization, the probability to choose the attack decision equals $p(A|G) = L^T M_A T \phi_G$; following a bad guy categorization, the probability to choose "attack" equals $p(A|B) = L^T M_A T \phi_B$. On D-alone trials, the probability to attack equals $p(A) = L^T M_A T \phi_I = L^T M_A T (p(G)\phi_G + p(B)\phi_B) = p(G) L^T M_A T \phi_G + p(B) L^T M_A T \phi_B = p(G) p(A|G) + p(B) p(A|B)$, and the latter equals the total probability to attack on C-D trials. This completes the proof. Note that this proof requires the same transition matrix T to be applied on both C-D as well as D-alone trials.

Appendix B

To begin, both models use a representation that has four *basis states* $\{GA, GW, BA, BW\}$, where, for example, GW symbolizes the combined event of categorizing the face as a good guy and deciding to withdraw. Evaluation of the payoffs causes the Markov state ϕ to be transformed by a transition matrix T into a decision state $\phi_F = T \cdot \phi$ used to make a choice about attacking or withdrawing. The transition matrix is defined by the following matrix exponential function (based on Busemeyer et al., 2009):

$$T = exp(K)$$

$$K = \begin{bmatrix} -1 & \frac{1-R}{R} k_G & 0 & 0 \\ 1 & -\frac{1-R}{R} k_G & 0 & 0 \\ 0 & 0 & -1 & \frac{R}{1-R} k_B \\ 0 & 0 & 1 & -\frac{R}{1-R} k_B \end{bmatrix},$$

The upper left corner of K is defined by the utility for attacking when the face is categorized as good; and the bottom right corner of K is defined by the utility for attacking when the face is categorized as bad.

Evaluation of the payoffs causes the quantum state ψ to be transformed by a unitary matrix U into a decision state $\psi_F = U \cdot \psi$ used to make a choice about attacking or withdrawing. The unitary

matrix is defined by the following matrix exponential function (based on Busemeyer et al., 2009):

$$U = exp\left(-i \cdot \frac{\pi}{2} \cdot (H_1 + H_2)\right)$$

$$H_1 = \begin{bmatrix} \frac{R\mu_G}{\sqrt{1+(R\mu_G)^2}} & \frac{1}{\sqrt{1+(R\mu_G)^2}} & 0 & 0 \\ \frac{1}{\sqrt{1+(R\mu_G)^2}} & \frac{-R\mu_G}{\sqrt{1+(R\mu_G)^2}} & 0 & 0 \\ 0 & 0 & \frac{R\mu_B}{\sqrt{1+(R\mu_B)^2}} & \frac{1}{\sqrt{1+(R\mu_B)^2}} \\ 0 & 0 & \frac{1}{\sqrt{1+(R\mu_B)^2}} & \frac{-R\mu_B}{\sqrt{1+(R\mu_B)^2}} \end{bmatrix},$$

$$H_2 = \frac{\gamma}{\sqrt{2}} \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

The upper left corner of H_1 is defined by the utility for attacking when the face is categorized as good; and the bottom right corner of H_1 is defined by the utility for attacking when the face is categorized as bad. The matrix H_2 aligns beliefs with actions of the decision-maker by amplifying the potentials for states GW, BA and attenuating potentials for states GA, BW to produce what is called an entanglement state (see Busemeyer & Bruza, 2012, chap. 9; Pothos & Busemeyer, 2009). For example, if $H_1 = 0$, then $\gamma = 1$ produces $|\psi_{GW}|^2 = .45 = |\psi_{BA}|^2$ and $|\psi_{GA}|^2 = .05 = |\psi_{BW}|^2$, producing an entangled state containing mainly contributions from GW, BA. The MATLAB program for computing predictions is available from the authors.

The following parameters were used to generate predictions from the models for Experiments 2 and 3.

Five Markov model parameters

p_c	$k_{G,b}$	$k_{B,b}$	$k_{G,g}$	$k_{B,g}$
.77	1.2901	.8448	.7291	.5508

Five Quantum model parameters (*forced, not free).

p_c	$\mu_{G,b}$	$\mu_{B,b}$	$\mu_{G,g}$	$\mu_{B,g}$*	γ
.79	−.1244	.3140	−.2729	.2729	.8500

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.01.019.

References

Aerts, D., Gabora, L., & Sozzo, S. (2013). Concepts and their dynamics: A quantum theoretic modeling of human thought. *Topics in Cognitive Science, 5*, 737–772.

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

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

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

Blutner, R., Pothos, E. M., & Bruza, P. (2013). A quantum probability perspective on borderline vagueness. *Topics in Cognitive Science, 5*, 711–736.

Brainerd, C. J., Wang, Z., & Reyna, V. (2013). Superposition of episodic memories: Overdistribution and quantum models. *Topics in Cognitive Science, 5*, 773–799.

Brainerd, C. J., Wang, Z., Reyna, V. F., & Nakamura, K. (2015). Episodic memory does not add up: Verbatim-gist superposition predicts violations of the additive law of probability. *Journal of Memory and Language, 84*, 224–245.

Bruza, P. D., Kitto, K., Nelson, D., & McEvoy, C. (2009). Is there something quantum like in the human mental lexicon? *Journal of Mathematical Psychology, 53*, 362–377.

Bruza, P. D., Kitto, K., Ramm, B. J., & Sitbon, L. (2015a). A probabilistic framework for analysing the compositionality of conceptual combinations. *Journal of Mathematical Psychology, 67*, 26–38.

Bruza, P. D., Wang, Z., & Busemeyer, J. R. (2015b). Quantum cognition: A new theoretical approach to psychology. *Trends in Cognitive Science, 19*(7), 383–393.

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

Busemeyer, J. R., & Pothos, E. M. (2012). Social projection and a quantum approach for behavior in prisoner's dilemma. *Psychological Inquiry, 23*, 28–34.

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

Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44*, 171–189.

Busemeyer, J. R., & Wang, Z. (2015). What is quantum cognition, and how is it applied to psychology? *Current Directions in Psychological Science, 24*(3), 163–169.

Busemeyer, J. R., Wang, Z., Khrennikov, A., & Basieva, I. (2014). Applying quantum principles to psychology. *Physica Scripta, T163*, 014007.

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

Busemeyer, J. R., Wang, Z., & Pothos, E. (2015a). Quantum models of cognition and decision. In J. R. Busemeyer, Z. Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 369–389). NY: Oxford University Press.

Busemeyer, J. R., Wang, Z., Pothos, E. M., & Trueblood, J. S. (2015b). The conjunction fallacy, confirmation, and quantum theory: Comment on Tentori, Crupi, and Russo (2013). *Journal of Experimental Psychology: General, 144*(1), 236–243.

Busemeyer, J. R., Wang, Z., & Shiffrin, R. (2015c). Bayesian comparison of a quantum versus a traditional model of human decision making. *Decision, 2*, 1–12.

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

Chaigneau, S. R., Barsalou, L. W., & Sloman, S. (2004). Assessing the causal structure of function. *Journal of Experimental Psychology: General, 133*, 601–625.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18.

Di Nunzio, G. M., Bruza, P., & Sitbon, L. (2014). Interference in text categorization experiments. In H. Atmanspacher, E. Haven, K. Kitto, & D. Raine (Eds.), *Quantum interaction. Lecture notes in computer science* (Vol. 8369, pp. 22–33). Springer.

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The SAGE handbook of prejudice, stereotyping and discrimination* (pp. 3–28). Los Angeles, CA: Sage.

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

Griffiths, O., Hayes, B. K., & Newell, B. R. (2012). Feature-based versus category based induction with uncertain categories. *Journal of Experimental Psychology: Learning, Memory, Cognition, 38*, 576–595.

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

Khrennikov, A. Y., & Haven, E. (2009). Quantum mechanics and violations of the sure thing principle: The use of probability interference and other concepts. *Journal of Mathematical Psychology, 53*, 378–388.

Krueger, J. I., DiDonato, T. E., & Freestone, D. (2012). Social projection can solve social dilemmas. *Psychological Inquiry, 23*, 1–27.

Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence. *Proceedings of the National Academy of Sciences, 112*(34), 10645–10650.

Maddox, T. W., & Bohil, J. B. (1998). Base rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning Memory and Cognition, 24*(6), 1459–1482.

Murdoch, D. (1987). *Niels Bohr's philosophy of physics*. Cambridge, UK: Cambridge University Press.

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology, 27*(2), 148–193.

Nosofksy, R. (2015). An exemplar-model account of feature inference from uncertain categorizations. *Journal of Experimental Psychology: Learning, Memory, Cognition*.

Pais, A. (1991). *Niels Bohr's times: In physics, philosophy and polity*. Oxford, NY: Clarendon Press.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability model explanation for violations of 'rational' decision making. *Proceedings of the Royal Society (B), 276*(1665), 2171–2178.

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences, 36*(3), 255–274.

Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological Review, 120*, 679–696.

Pothos, E. M., & Trueblood, J. S. (2014). Structured representations in a quantum probability model of similarity. *Journal of Mathematical Psychology, 64*(C), 35–43.
Rehder, B., & Burnette, R. C. (2005). Feature inference and the causal structure of object categories. *Cognitive Psychology, 50*, 264–314.
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.
Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science, 35*, 1518–1552.
Trueblood, J. S., & Busemeyer, J. R. (2012). A quantum probability model of causal reasoning. *Frontiers in Cognitive Science, 3*, 1–13.
Wang, Z., & Busemeyer, J. R. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science, 5*, 689–710.
Wang, Z., & Busemeyer, J. R. (2015). Reintroducing the concept of complementarity into psychology. *Frontiers in Psychology, 6*. Article 1822.
Wang, Z., & Busemeyer, J. R. (2016). Order effects in sequential judgments and decisions. In H. Atmanspacher & S. Maasen (Eds.), *Reproducibility: Principles, practices, and problems*. San Francisco, CA: Wiley.
Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science, 5*, 672–688.
Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences, 111*(26), 9431–9436.
White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). Sometimes it does hurt to ask: The constructive role of articulating impressions. *Cognition, 133*(1), 48–64.
Yukalov, V. I., & Sornette, D. (2011). Decision theory with prospect interference and entanglement. *Theory and Decision, 70*(3), 283–328.