Decision
2015, Vol. 2, No. 1, 1–12

2325-9965/15/$12.00 http://dx.doi.org/10.1037/dec0000017

Bayesian Model Comparison Favors Quantum Over Standard Decision Theory Account of Dynamic Inconsistency

Jerome R. Busemeyer
Indiana University

Zheng Wang
The Ohio State University

Richard M. Shiffrin
Indiana University

Many paradoxical findings in decision-making that have resisted explanations by standard decision theories have accumulated over the past 50 years. Recent advances based on quantum probability theory have successfully accounted for many of these puzzling findings. Critics, however, claim that quantum probability theory is less constrained than standard probability theory, and hence quantum models only fit better because they are more complex than standard decision models. In this article, for the first time, a Bayesian method was used to quantitatively compare the 2 types of decision models, which is a method that evaluates models with respect to accuracy, parsimony, and robustness. A large experiment was used to compare the best-known models of each type, matching in their numbers of parameters, but possibly differing in the complexity of their functional forms. Surprisingly, the Bayesian model comparison overwhelmingly favored the quantum model, indicating that its success is due to its robust ability to make accurate predictions rather than accidental fits afforded by increased complexity.

Keywords: quantum, dynamic inconsistency, Bayesian model comparison, two-stage gamble, prospect theory

Supplemental materials: http://dx.doi.org/10.1037/dec0000017.supp

Recently, a new theoretical framework for constructing models of human judgment and decision-making has been proposed, which is based on mathematical principles derived from quantum probability theory (Busemeyer & Bruza, 2012; Khrennikov, 2010; Pothos & Busemeyer, 2012; Wang, Busemeyer, Atmanspacher, & Pothos, 2013). This new framework does not rely on the assumption that the brain is some kind of quantum computer. Instead, it uses a probabilistic formulation from quantum theory, which is built on a noncommutative algebra of operators, to explain human decision-making behavior. Quantum decision theory has successfully accounted for stubborn problems that have resisted coherent explanations by conventional decision theories for decades, including violations of Savage's "sure thing" principle of rational decision-making (Pothos & Busemeyer, 2009), conjunction and disjunction probability judgment fallacies (Busemeyer et al., 2011), over- and under- extension errors in conceptual reasoning (Aerts, 2009), asymmetric similarity judgments (Pothos et al., 2013), order effects on probabilistic inference (Trueblood & Busemeyer, 2011), interference of categorization on decision-making (Busemeyer et al., 2009), attitude question order effects (Wang & Busemeyer, 2013), and other puzzling results from decision research (Lambert-Mogiliansky et al., 2009; La Mura, 2009; Yukalov & Sornette, 2011). In short, quantum decision models have made impressive progress organizing and ac-

This article was published Online First November 3, 2014.

Jerome R. Busemeyer, Department of Psychological and Brain Sciences, Indiana University; Zheng Wang, Department of Communication, The Ohio State University; Richard M. Shiffrin, Department of Psychological and Brain Sciences, Indiana University.

This research was supported by NSF SES-1153726, SES-1153846, and AFOSR FA 9550-12-1-0397.

Correspondence concerning this article should be addressed to Jerome R. Busemeyer, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405. E-mail: decisionjrb@gmail.com

counting for a wide range of perplexing findings in human judgment and decision-making using a common set of axiomatic principles.

Quantum probability theory relaxes some constraints of traditional probability theory, and so it represents a type of generalized probability theory (Gudder, 1988). However, this increase in generality raises the question: Does the success of quantum models rely on greater model complexity? Although this question is often raised (Rakow, 2013; Behme, 2013; Shanteau & Weiss, 2013), it has rarely been formally examined. In one case where it was formally examined (Atmanspacher & Roemer, 2012), the quantum probability model was less complex than a traditional probability model. However, this remains an important question that deserves further consideration. The purpose of this article is to rigorously compare quantum versus standard decision models using Bayesian model selection, which is a state-of-the art method for comparing models according to their accuracy, parsimony, and robustness (Lee & Wagenmakers, 2014).

The competing models are compared with respect to their predictions for a large decision-making experiment conducted by Barkan and Busemeyer (2003), which was designed to examine a rational principle called dynamic consistency. According to optimal theory, when faced with decisions involving sequences of actions and events across time, the decision-maker should work backward by forming optimal plans for future decisions in order to determine the best course of action for the current decision. A decision-maker is dynamically consistent if the person actually carries out the planned action once that decision is reached (Machina, 1989; Sarin & Wakker, 1998). Changing the plan constitutes a violation of dynamic consistency, which leads to suboptimal strategies. Nevertheless, violations of dynamic consistency have been observed in simple risky decision tasks (Barkan & Busemeyer, 1999, 2003; Busemeyer et al., 2000; Cubitt et al., 1998; Hey & Knoll, 2007).

The experiment by Barkan and Busemeyer (2003) was selected for the model comparison for several reasons. First, the dynamic inconsistency findings from this experiment provide an interesting challenge for decision theories to explain. Second, this experiment used a large within-subjects design with a wide range of 17 payoff conditions, which affords model comparisons at the level of individual participants. Third, the experiment included a large sample of 100 participants for evaluating the generality of the model comparison results. Fourth, a standard decision model, based on the popular prospect theory (Tversky & Shafir, 1992) has already been developed and shown to provide accurate fits to the human behavioral data. Thus the quantum model is pitted against the best known and well-performing standard model for these empirical findings.

Here we review the experimental methods and basic findings, describe the two decision models that are compared, present the main findings from the Bayesian model comparison analysis, and then draw our conclusions. However, before these, first we justify our use of the Bayesian method for model comparison.

Bayesian Model Comparison

A Bayesian model comparison method was used to compare the quantum versus traditional decision models for several reasons. First and foremost, Bayesian model selection is based on rational principles of inference rather than ad hoc rules. That is, Bayes's theorem is used to select the model with the greatest posterior probability computed from the strength of all of the evidence supporting each model (Kass & Raftery, 1995; Jaynes, 2003). The Bayesian updating rule provides a coherent principle for accumulating evidence for each model across replications, which is a crucial issue for the present work. A single participant in the empirical data included in the analysis may provide only modest support for one model over another, and stronger evidence for or against a model must be obtained by accumulating the evidence across all participants in a rational manner. Second, Bayesian methods can be applied to non-nested models, and frequentist chi-square tests of differences between models can only be used with nested models. The models in the present application are non-nested, because the range of predictions generated by the quantum model cannot be computed from the traditional model and visa versa. Third, the Bayes factor is affected by not only the number of parameters, but also by model flexibility and ranges of parameter spaces. Commonly used model comparison methods, such as Akaike information criterion or Schwartz's Bayesian information criterion, only penalize for the number of model parameters. In the present application, both models use the same number of parameters, but this does not mean they have the same model complexity or parsimony. It is well known that model complexity is more than just the number of parameters, and models with the same number of parameters can vary with respect to model flexibility (Myung & Pitt, 1997; Jeffreys & Berger, 1992). Finally, the Bayes factor compares how *robustly* each model accurately predicts the data over the plausible range of parameter values (Shiffrin et al., 2008). This point is elaborated next.

Bayesian model comparison is based on the calculation of each model's expected likelihood, given the data. Each expectation is obtained by a weighted average of all of the likelihoods produced by a model over the *entire* model parameter space, using weights defined by a prior distribution on the model parameter values. The ratio of the expected likelihoods of the competing models forms the *Bayes factor*. Unlike frequentist methods, no parameters are ever "fit" to the data using the Bayesian method. Instead, the parameters for each model are selected a priori, and the likelihood of the data produced by these a priori values are evaluated. However, when computing the Bayes factor, the evaluation includes all of the plausible parameter values as determined by the prior distribution. A model is robust if it produces a high likelihood across a wide range of its plausible parameter values. Overly complex models are not robust, because they only produce high likelihoods fortuitously, and only within a small region of the plausible parameter space. Models that are not robust are penalized by averaging the likelihood over the entire range of the plausible parameter space, including plausible regions producing low likelihoods. Robust models are rewarded by averaging the likelihood over the entire range of the plausible parameter space, because the likelihood remains relatively high. Thus, the Bayes factor favors the model that robustly makes accurate predictions across the entire range of its plausible parameter values. (Shiffrin et al., 2008).[1]

Of course, like any other method, the Bayesian method has its drawbacks. In particular, we must specify a priori distribution over parameters. To a large extent this is based on experience from previous model fits. But this issue can be mitigated by examining more than one priori, as we do in this study. Nevertheless, it is generally useful to consider multiple methods for model comparison. In the SI, we report results from other model comparison methods based on least squares fits (using *R*-square criteria) to means averaged across participants, as well as maximum likelihood methods (using Akaike and Bayesian information criteria) applied to individual participants. To briefly summarize the results reported in the Supplemental Materials, all these alternative methods favored the quantum over the standard model. (The two models have the same number of parameters.)

The Two-Stage Gambling Task and Dynamic Inconsistency

Tversky and Shafir (1992) originally invented a two-stage gambling task to investigate Savage's "sure thing" principle Savage (1954). According to this principle, if under the state of the world X, you prefer action A over B, and if under the complementary state of the world not-X, you also prefer action A over B, then you should prefer A over B even when the state of the world is unknown. The two-stage gambling task was used to test this principle as follows. The participant is required to play a gamble (with equal chance to win \$200 or lose \$100) during the first stage, and then this person is given a choice whether or not to play the same gamble again during the second stage. This choice is made under three different conditions: assuming the person (a) has won the first play (the "known win" condition), (b) has lost the first play (the "known loss" condition), (c) does not know the result of the first play (the "unknown" condition). The results of this initial study showed that participants generally preferred to play the gamble again for both the known win (.69 take the gamble) and known loss (.59 take the gamble) conditions, but they generally preferred not to play under the unknown condition (.39 take gamble). These findings indicate a violation of the "sure thing" principle. Tversky and Shafir proposed a model

[1] The validity of using Bayes factor (based on traditional probability) for comparing models (one of which is based on quantum probability) is explained in the Supplemental Materials.

based on prospect theory to account for these results, which is used to formalize the standard model in our model comparison.

The two-stage gambling paradigm was modified by Barkan and Busemeyer (1999, 2003) to investigate dynamic consistency. Participants were required to play a gamble (with equal chance to win x or lose y) in the first stage, and then they were asked to make two decisions, a plan and a final decision, about the second stage for the same gamble. The plan for the second stage gamble was made before learning the outcome of the first stage gamble ("Do you plan to play again if you win/lose the first time?"). The final decision was made after learning the outcome of the first stage (e.g., "After actually winning the first time, do you now want to play again?"). Half of the trials were randomly assigned to a win on each stage, and the remaining half were assigned to a loss. (Participants observed a computer simulated coin flip to discover the first stage outcome.) Real monetary payments (such as those shown in Table 1) were made at the end of the experiment by randomly picking out four trials, randomly sampling either the plan or the final decision to determine the choice, and selecting the payoffs assigned to the choices for the selected trials.

Although the larger study by Barkan and Busemeyer (2003) was used for model comparison, Table 1, taken from Barkan and Busemeyer (1999), provides a succinct summary of the basic results. The first pair of columns under the label "Gamble" show the payoffs (in units of dollars) of a gamble; the next pair of columns under the labels "Plan Win" and "Plan Loss" show the proportion of taking the second stage gamble when planning on winning or losing the first stage; the third pair of columns under the labels "Final Win" and "Final Loss" show the proportion of taking the gamble during the second stage after learning about winning or losing the first stage gamble. As can be seen in the table, participants were dynamically inconsistent and changed their plans: following a win, participants tended to change from planning to take toward actually rejecting the gamble in the final stage; following a loss, participants tended to change from planning to reject toward actually taking the gamble in the final stage. These findings were replicated and extended in the larger study by Barkan and Busemeyer (2003) that included 100 participants each making plans and final decisions for 33 gambles formed from 17 payoffs conditions.[2] Once again, this larger study is used for the model comparison.

Table 1
Summary of Results From Barkan and Busemeyer (1999)

Gamble		Choice proportions*			
Win	Loss	Plan win	Plan loss	Final win	Final loss
\$0.80	\$1.00	.25	.26	.20	.35
\$0.80	\$0.40	.76	.72	.69	.73
\$2.00	\$1.00	.68	.68	.60	.75
\$2.00	\$0.40	.84	.86	.76	.89

* Each choice proportion is based on four replications by 100 participants. Note that Barkan and Busemeyer (2003) was used for the model comparison.

Choice Models for the Two-Stage Gambling Task

Reference Point Model (Model R)

Barkan and Busemeyer (2003) accounted for their results by using a model based on prospect theory (Tversky & Kahneman, 1990), which was earlier used by Tversky and Shafir (1992) to account for the violation of the "sure thing" principle. The essential idea is that the decision-maker ignores first stage outcome during the plan, but then later incorporates this outcome during the final decision, which causes a change in the reference point of the utility function used for the plan versus the final decisions.

Consider a generic gamble G that produces (with equal probability) either a win equal to x_W or a loss magnitude equal to x_L. One commonly used utility function for a payoff x is a power function that includes risk aversion and loss aversion: $u(x) = x^a$ *for* $x \geq 0$, and $u(x) = -b \times |x|^a$ *for* $x < 0$ (hereafter, we call this the conventional utility function). The parameter a is used to model risk aversion and it usually has a value below 1; the parameter b is used to model loss aversion and usually it has a value greater than 1. A less frequently used utility function is a pair of power functions: $u(x) = x^a$ for $x \geq 0$

[2] Sixteen of the payoff conditions were replicated twice; one payoff condition was presented once. See Supplemental Materials for a summary of these results.

and $u(x) = -|x|^b$ *for* $x < 0$ (hereafter, we will call this the alternative utility function). Even though the outcomes are all equally likely, we can also apply a different decision weight $0 \leq w_G \leq 1$ for winning and $w_L = 1 - w_G$ for losing the gamble (hereafter, we will call this the unequal decision weight model).

During planning, it is assumed that people ignore the unknown first stage outcome and simply compute a utility for playing the second gamble based solely on the payoffs of the second gamble:

$$u(G\,|\,\text{Plan}) = w_G \cdot u(x_W) + w_L \cdot u(-x_L), \quad (1)$$

The choice for the plan is based on the comparison of the utility of gambling on the second play to the status quo (a zero outcome), $D_P = u(G|\text{Plan}) - 0$. For example, using the conventional utility function with coefficients $a = .65$, $b = 1.6$, $w_G = w_L = .50$, and setting $x_W = \$2$ and $x_L = \$1$, we obtain $D_P = -.0154$, which implies that the decision-maker prefers not to take the gamble again.

After experiencing a win, the person includes the win from the first gamble into the evaluation of the second gamble's payoffs by using the following utility function:

$$u(G\,|\,\text{Win}) = w_G \cdot u(x_W + x_W) + w_L \cdot u(x_W - x_L) \quad (2)$$

Then the choice is based on the comparison of the utility of gambling again on the second play to the utility of keeping the win from the first gamble, $D_W = u(G\,|\,Win) - u(x_W)$. Continuing with the previous example, using the conventional utility function and setting $a = .65$, $b = 1.6$, $w_G = w_L = .50$, and setting $x_W = \$2$ and $x_L = \$1$, we obtain $D_W = .162$, which implies that the decision-maker now prefers to take the gamble.

Similarly, after experiencing a loss, the person includes the loss from the first gamble into the evaluation the second gamble's payoff by using the following utility function:

$$u(G\,|\,\text{Loss}) = w_G \cdot u(-x_L + x_W) + w_L \cdot u(-x_L - x_L) \quad (3)$$

Then the choice is based on the comparison of the utility of gambling again on the second play to the utility of keeping the loss from the first gamble, $D_L = u(G\,|\,Loss) - u(x_L)$. Continuing with the previous example, setting $a = .65$, $b = 1.6$, $w_G = w_L = .50$, and setting $x_W = \$2$ and $x_L = \$1$, we obtain $D_L = .8447$, which again implies that the decision-maker prefers to take the gamble.

Essentially, the reference point for evaluating gains and losses changes in this model during the plan versus the final decisions. For example, during the plan, the possibility of losing \$1 on the second gamble is evaluated as a loss (because any payoff below zero is considered a loss). However, after finding out that \$2 was won on the first play, the possibility of losing \$1 on the second play is still evaluated as an overall gain (any payoff below –\$2 is now considered a loss). In short, dynamic inconsistency arises from the use of different utility functions, which are defined by different reference points, for the plan versus the final decisions.[3]

So far, this model is deterministic. To convert utilities into probabilities, it is common to assume an extreme value random utility model (McFadden, 1981). Under this assumption, the choice probabilities are determined by a logistic probability distribution function, which maps the utility difference D_j for j = {Win, Lose, Unknown} into probabilities by the function $p(T\,|\,j) = 1/(1 + e^{-\gamma \cdot D_j})$, where the parameter γ adjusts the sensitivity of choice probability to the gambles utility. For example, if $D_j = 1.0$, and $\gamma = 2$, then $p(T\,|\,j) = .88$.

This model has four parameters (a, b, w_G, γ) representing risk aversion, loss aversion, decision weight for gain, and the choice probability parameter. Model R does a reasonably good job of fitting the 17 (gambles) × 2 (plan vs. final) = 34 mean choice proportions reported in Barkan and Busemeyer (2003). Using only three parameters (setting $w_G = w_L$ because unequal weights did not improve fit) produced a percentage of predicted variance equal to $R^2 = .77$ (for both conventional and alternative utility functions).

[3] Barkan and Busemeyer (2003) also compared the reference point version of prospect theory described here to another model that assumed experience changed the subjective probabilities of winning and losing (instead of changing the reference points), but the reference point model fit 68% of the individuals better. Hence, the reference point model is selected here as the competing model.

Quantum Model (Model Q)

The quantum model of the dynamic inconsistency effect is the same model that has previously been used to account for violations of the aforementioned "sure thing" principle by Pothos and Busemeyer (2009). Unlike the reference point model that uses inconsistent utility functions to account for dynamic inconsistency, the quantum model assumes that decision-makers use a consistent utility function for both plan and final choices, which always incorporates the outcomes from the first stage. Instead, dynamic inconsistency arises in the quantum model because there is *uncertainty* about the first stage outcome during the plan, but this *uncertainty* is resolved during the final stage. According to the quantum model, the resolution of uncertainty produces an interference effect that manifests itself as dynamic inconsistency (Yukalov & Sornette, 2009).

In general, quantum probability theory uses a vector space to represent events, and a state vector within this space to determine probabilities of events. The two-stage game involves a set of four combinations of events $\{WT, WR, LT, LR\}$. For example, WT symbolizes the combination "win the first stage gamble" and "take the second stage gamble," and LR represents the combination "lose the first stage gamble" and "reject the second stage gamble." These four combinations of events are represented by a four-dimensional vector space spanned by four basis vectors symbolized as $\{|WT\rangle, |WR\rangle, |LT\rangle, |LR\rangle\}$. The four basis vectors are assumed to be orthogonal because the four combinations they represent are mutually exclusive. The decision-maker has a belief about each event (win or lose the first play) and a potential to take each action (take or not take the second gamble). These beliefs and action potentials are represented by a state vector $|\psi\rangle$ within the four-dimensional vector space. For example, if the person is in state $|\psi\rangle = |WT\rangle$, then the person is certain that a win occurred on the first stage and the person is certain to take the second gamble; alternatively if the person is in state $|\psi\rangle = |LR\rangle$, then the person is certain that a loss occurred on the first stage and the person is certain to reject the second gamble. However, the person's state does not have to be exactly aligned with one of the basis vectors, and instead the person can be in a *superposition* state $|\psi\rangle = \psi_{WT}|WT\rangle + \psi_{WR}|WR\rangle + \psi_{LT}|LT\rangle + \psi_{LR}|LR\rangle$. In this case, $|\psi_{WT}|^2$ represents the probability that the person believes a win occurred on the first stage and decides to take the second gamble; likewise $|\psi_{LR}|^2$ represents the probability that the person believes that a loss occurred on the first stage and decides to to reject the second gamble. The state vector $|\psi\rangle$ is assumed to be unit length $|||\psi\rangle|| = 1$ so that the squared magnitudes of the four coordinates sum to one. The superposition state is a unique concept in quantum theory—it represents the idea that priori to decision, the decision-maker is indefinite with respect to the four basis states, and each possible basis state has some potential to become expressed at the moment of decision. In this application, when the outcome of the first play is unknown, the person is superposed between thinking that she won or lost the first play, and he or she is also superposed between deciding to take or not take the second gamble. In our study, this superposition state is affected by information about the first stage outcomes as well as payoffs for taking or rejecting the gamble. For convenience, we use a 4×1 column matrix ψ to represent the tetrad $(\psi_{WT}, \psi_{WR}, \psi_{LT}, \psi_{LR})$ of coordinates (also called amplitudes) assigned to the state vector $|\psi\rangle$.

The initial state is represented by ψ_0, which serves as a "prior" amplitude distribution over the four basis states. At this point, the outcome of the first stage is unknown and the payoffs have not been evaluated, and so a uniform prior distribution is assumed so that ψ_0 has elements $\psi_{ij} = 1/2$ for all four entries. This superposition state represents the *uncertainty* that exists about the first stage game during the plan before evaluating payoffs. (Note that the squared length of this vector equals one as required.)

The *uncertainty* about the first stage game is resolved after learning the first stage outcome, and consequently, the decision-maker's state changes to a new information state ψ_1, which depends on the type of information provided about the first stage. The state following experience of a win is updated to $\psi_1 = \psi_W$, which has $1/\sqrt{2}$ in the first two entries (consistent with a first stage win) and zeros in the second two (inconsistent with a first stage win). The state following experience of a loss is updated to $\psi_1 = \psi_L$, which has $1/\sqrt{2}$ in the last two entries (consistent with a first stage loss) and zeros in the first two entries (inconsistent with a first stage loss). This is the state after learning about the first stage outcome but before evaluating the payoffs. (Note that the squared length of each of these vectors equals one as required.)

Before taking an action, the decision-maker needs to evaluate the expected payoffs produced by the gamble. Recall that before this evaluation, the decision-maker is in a state ψ (either ψ_0 for the planning stage or ψ_1 after learning the first stage outcome) that is equally likely to take or not take the gamble. Evaluation of the payoffs is achieved by rotating the decision-maker's state ψ toward or away from taking the gamble (without changing its length). Mathematically, this rotation is represented by a 4×4 unitary matrix U that transforms the state ψ into a decision state $\psi_D = U \cdot \psi$.

The unitary matrix U is a function of $D_W = u(G|Win) - u(x_W)$ as defined by Equation 2, and $D_L = u(G|Loss) - u(x_L)$ as defined by Equation 3. If a win occurred on the first stage, then only D_W affects the rotation; if a loss occurred on the first stage, then only D_L affects the rotation, but if the first stage outcome is unknown, then both affect the rotation. (The mathematical details for constructing the unitary matrix U are described in the Supplemental Materials; also see Pothos & Busemeyer, 2009, or Busemeyer & Bruza, 2012, Ch. 9.) Because the unitary matrix U depends on the gamble, it varies across gambles.

The unitary matrix U is also a function of a parameter γ that is used to align beliefs with actions. A single value for γ is used for all gambles. This alignment is achieved by amplifying the potentials for states *WT*, *LR* and attenuating potentials for states *WR*, *LT* to produce what is called an "entangled" state. In particular, this entanglement captures the idea of a belief in a "hot hand," for example, winning the first hand is correlated with playing again on the second hand. (Once again, see the Supplemental Materials, or Pothos & Busemeyer, 2009, or Busemeyer & Bruza, 2012, Ch. 9 for details).

A projection matrix $M = diag[1,0,1,0]$ is used to map the decision state into response probabilities for taking the second stage gamble:

$$p(T) = \|M \cdot \psi_D\|^2 = \|M \cdot U \cdot \psi\|^2, \quad (4)$$

More specifically, $p(T|\text{Plan}) = \|M \cdot U \cdot \psi_0\|^2$, $p(T|\text{Win}) = \|M \cdot U \cdot \psi_W\|^2$, and $p(T|\text{Loss}) = \|M \cdot U \cdot \psi_L\|^2$. For example, if we use the conventional utility function and set $a = 1$, $b = 1$, $w_G = .75$, $\gamma = 1.74$, and setting $x_W = \$2$ and $x_L = \$1$, then $p(T|Win) = .70$ and $p(T|Loss) = .56$ and $p(T|Plan) = .38$, which closely matches the results found in Tversky and Shafir (1992).

The inconsistency between the plan and final decision occurs as follows:

$$\begin{aligned} p(T|\text{Plan}) &= \|M \cdot U \cdot \psi_0\|^2 \\ &= \|M \cdot U \cdot \tfrac{1}{\sqrt{2}}(\psi_W + \psi_L)\|^2 \\ &= \tfrac{1}{2}\|M \cdot U \cdot \psi_W + M \cdot U \cdot \psi_L\|^2 \\ &= \tfrac{1}{2}p(T|\text{Win}) + \tfrac{1}{2}p(T|\text{Loss}) + Int, \end{aligned}$$

where *Int* represents the sum of the cross product terms produced by squaring the sum, which can be positive or negative. As can be seen here, the probability of taking the gamble during the plan is an average of the two final probabilities plus interference. The interference term depends on the entanglement parameter γ, and if $\gamma = 0$, then the interference term equals zero. It is clear from Table 1 that the interference cannot be zero, because the choice probabilities for the plan do not equal the average of the two final choice probabilities.

This model has four parameters (a, b, w_G, γ) and only the last one differs in interpretation from Model R. Model Q does a reasonably good job of fitting the 17 (gambles) $\times$ 2 (plan vs. final) = 34 mean choice proportions reported in Barkan and Busemeyer (2003). Using only three parameters (once again setting $w_G = w_L$) produced a percentage of predicted variance equal to $R^2 = .82$ and $R^2 = .85$ for conventional and alternative utility functions, respectively.

In summary, the two models differ in how they conceptually explain dynamic inconsistency: The reference point model uses inconsistent utility functions for the plan and the final decisions to produce dynamic inconsistency. According to the quantum model, inconsistency arises from a state of uncertainty that exists during the planning process which is resolved during the final choices. The models are next compared using a rigorous Bayesian model selection method at the individual level of analysis.

Model Comparisons Based on Bayes Factors

Log Likelihood Function

Next, we describe the log likelihood function for the data from each individual using the four combinations of plan and final choices observed from 33 gambles that each person completed. On each trial, a gamble was presented and the participant made both a plan (conditioned on an anticipated outcome) and a final choice (conditioned on observing that same outcome). For person i on trial t, we observe a data pattern $X_i(t) = [x_{TT}(t), x_{TR}(t), x_{RT}(t), x_{RR}(t)]$, defined by $x_{jk}(t) = 1$ if event (j, k) occurs and otherwise zero, where TT is the event "planned to take the gamble and finally took the gamble," TR is the event "planned to take the gamble but finally rejected the gamble," RT is the event "planned to reject the gamble but finally took the gamble," and RR is the event "planned to reject the gamble and finally rejected the gamble."

To allow for possible dependencies between a pair of choices within a single trial, an additional memory recall parameter was included in each model. For both models, it was assumed that there is some probability $0 \leq m \leq 1$ that the person simply recalls and repeats the planned choice during the final choice, and there is some probability $1 - m$ that the person forgets or ignores the planned choice when making the final choice. After including this memory parameter, the prediction for each event becomes

$$\begin{aligned} p_{TT} &= p(T \mid plan) \cdot (m \cdot 1) + (1 - m) \cdot p(T \mid final) \\ p_{TR} &= p(T \mid plan) \cdot (1 - m) \cdot p(R \mid final) \\ p_{RT} &= p(R \mid plan) \cdot (1 - m) \cdot p(T \mid final) \\ p_{RR} &= p(R \mid plan) \cdot (m \cdot 1) + (1 - m) \cdot p(R \mid final) \end{aligned}$$

Using these definitions for each model, the log likelihood function for the 33 trials (with a pair of choices on each trial) from a single person can be expressed as

$$\ln L(X_i(t)) = \sum_{j,k} x_{jk}(t) \cdot \ln (p_{jk})$$
$$\ln L(X_i) = \sum_{t=1}^{33} \ln L(X_i(t)).$$

To check how well each decision model fits the choice data for individuals, we compared each decision model to a null model and also to a saturated model. The decision models use only four parameters per person (except allowing unequal weigh adds one extra). The saturated model defines p_{jk} as the sample proportion observed for a gamble from an individual, and so this model has $(4 - 1) \times 17 = 51$ parameters per person. The null model set $p(T|plan) = p(T|final) = .50$, and so it used only a single memory parameter m per person. The statistic $G_i^2 = -2 \cdot lnL(X_i)$ (using maximum likelihood estimates) was computed for each model and person, and the total G^2 for a model was obtained by summing across the 100 participants. Chi-square tests were performed by comparing changes in total G^2 to chi square critical values (degrees of freedom equal to the difference in number of model parameters). The total G^2 equals 3,100 (saturated model), 6,883 (Model Q with conventional utility, equal weight), 6,936 (Model R conventional utility, unequal weight), 7,020 (Model R conventional utility, equal weight), 7,124 (Model R alternative utility, equal weight), and 7,688 (null model). The difference between the null model versus each decision model was statistically significant ($p < .05$); there were no statistically significant differences between the saturated model and each decision model for all participants (see Supplemental Materials for more details).

Prior Distributions

Each model has five parameters $\theta = (a, b, w_G, \gamma, m)$: a risk aversion parameter, a loss aversion parameter, a decision weight parameter, a choice model parameter, and a memory parameter. Four parameters are common across both models, and they only differ with respect to a fifth parameter. Given how little we know about the likelihood function of the quantum model, we used a very conservative method to determine this function across the parameter space. We used a very fine grid of points per parameter. The number of grid points was chosen as follows. We refined the grid until the change in $logBF$ for each person converged to less than .01. As shown in the Supplemental Materials, this criterion was already achieved for all participants using only 21 grid points, but to guarantee convergence, we increased the size to 41 grid points. In fact, comparing results from grid of 41 points with a less-fine grid using only 21 points revealed no meaningful difference between the final Bayes factors. The 41-point grid generated $41^4 = 2,825,761$ combinations, and we evaluated the log likelihood function for each model at each combination. The range of parameter values chosen for each parameter was based on a priori values of the standard decision models:

$$a \in [.400, .425, \ldots, .875, \mathbf{.90}, .925, \ldots, 1.375, 1.40],$$
$$b \in [.50, .55, \ldots, 1.45, \mathbf{1.50}, 1.55, \ldots, 2.45, 2.50], \text{ loss aversion}$$
$$w_G \in [.25, .2625, \ldots, \mathbf{.50}, \ldots, .7375, .75],$$
$$m \in [.000, .025, \ldots, .475, \mathbf{.50}, .525, \ldots, .975, 1.00],$$
$$\gamma \in [0, .10, \ldots, 1.90, \mathbf{2.0}, 2.10, \ldots, 3.90, 4.0], \text{ (R)}$$
$$\gamma \in [-5.00, -4.75, \ldots, -.25, \mathbf{0.00}, .25, \ldots, 4.75, 5.00]. \text{ (Q)}$$

The risk aversion parameter a ranges from risk aversion to risk seeking centered at slight risk aversion. For the conventional utility, the loss aversion parameter function b ranges across loss insensitivity to loss sensitivity centered at slight loss aversion. For the alternative utility function, the same range was used for both b and a. The decision weight parameter w_G ranges from below to above .50 centered at .50. The memory parameter m ranges from no recall to perfect recall centered at .50. These ranges were used for both models. Only the choice parameter γ differs between the two models: It ranges from random to deterministic choice for the reference point model centered at a value commonly found in previous studies, and it ranges from positive to negative values for the quantum model.

Two different prior distributions were examined: a uniform and a (discretized) normal distribution. The uniform distribution assigned equal probability to each grid combination point. For the normal prior, we assumed independent (discretized) normal distributions with means equal to the center of each range and standard deviations that covered the range for each parameter (see Supplemental Materials for more details). The normal prior for the risk aversion parameter a was assigned a mean of .90 (slight risk aversion) and a standard deviation of .75. For the conventional utility function, the normal prior for the loss aversion parameter b was assigned a mean of 1.5 (slight loss aversion) and a standard deviation equal to .75. For the alternative utility model, the normal prior used to define parameter b was exactly the same as used for parameter a. The decision weight parameter was normally distributed around a mean of .50 and a standard deviation equal to .75. The memory parameter was normally distributed around a mean of .50 and a standard deviation equal to .75. The prior of the gamma parameter for the reference point model was normally distributed around a mean of 2 and a standard deviation equal to 2.5. The prior of the gamma parameter for the quantum model was normally distributed around a mean of zero with a standard deviation equal to 10.

Bayes Factor

The Bayes factor was computed for each person and each model by first computing the expected likelihood for each model for a person, $p_M(X_i)$, which is a weighted average of all the likelihoods. For each combination of parameter values, we compute a likelihood $L_M(X_i \mid \theta)$ and assign a prior probability $w(\theta)$. The Bayes factor (denoted BF) for each person equals the ratio of the expected likelihoods of the competing models.

$$p_M(X_i) = E[L_M(X_i \mid \theta)] = \sum_{\theta} w(\theta) \cdot L_M(X_i \mid \theta)$$

$$BF_i = \frac{p_{M1}(X_i)}{p_{M2}(X_i)},$$

where M1 and M2 are two models being compared. The Bayes factor ranges from 0 to infinity, with $BF < 1$ favoring Model M2 and $BF > 1$ favoring Model M1. The evidence scale becomes more symmetric by taking the natural log of the Bayes Factor, $lnBF_i$, in which case negative values favor the Model M2 and positive values favor Model M1. A log Bayes factor is obtained from each individual, and the accumulated evidence across all independent participants equals sum $lnBF = \sum_{i=1,100} lnBF_i$. The Bayes factor for the experiment actually equals the exponential of this sum.

First we compared Model R with equal ($w_G = w_L$) versus unequal decision weights ($w_G \neq w_L$) using the conventional utility function. The Bayes factor for this model comparison strongly favors the equal weight over the unequal weight version of Model R. Using the uniform prior distribution, we obtained sum $lnBF = 53.15$ in favor of the equal weight model, and 90% of the individuals produced positive log Bayes factors favoring equal weight; using the normal prior distribution we obtained $lnBF = 25.43$ and 69% of the participants produce positive log Bayes factors favoring equal weight. Hereafter, we will only use the equal weight model ($w_G = .50 = w_L$) for all comparisons (because using the unequal weight model will only reduce the Bayes factor for Model R).

Next we compared Model R versus Q (setting $w_W = w_L = .50$) using the conventional utility function. The Bayes factor for this model comparison strongly favors Model Q over Model R. Using the uniform prior distribution, we obtained sum $lnBF = 74.5$, favoring Model Q, and 90% of the participants produced positive log Bayes factors; using the normal prior distribution we obtained sum $lnBF = 83.05$, and 93% of the participants produced positive log Bayes factors. We also compared these models by reducing the range of the uniform distribution by one half, and by reducing the standard deviation for the normal distributions by one half, but the Bayes factor continued to strongly favor the quantum model (see Supplemental Materials for details).

Finally, we compared Model R versus Q (setting $w_W = w_L = .50$) using the alternative utility function. Once again, the Bayes factor favors Model Q, but the evidence was less strong for this case. Using the uniform prior distribution, we obtained sum $lnBF = 11.23$, favoring Model Q, and 63% of the participants produced positive log Bayes factors; using the normal prior distribution, we obtained sum $lnBF = 5.11$, favoring Model Q for the normal distribution, and 54% of the participants produced positive log Bayes factors. This last result indicates that part of the problem for the prospect model lies with the utility function used to produce the dynamic inconsistency effect.

Recall the Model R accounts for the dynamic inconsistency effect by changing the utility function for plan versus final decisions. Apparently the predictions of Model R for this data set are sensitive (not robust) with respect to this utility function. This lack of robustness is penalized by the Bayes factor. Alternatively, the quantum model accounts for the dynamic inconsistency effect by the interference effect produced by the reduction of uncertainty. The interference effect is moderated by the parameter γ and the predictions of Model Q are fairly robust with respect to this parameter (as shown in the Supplemental Materials, the likelihood function produces a damped oscillation symmetrically around zero, but it remains fairly high at all values except near zero). Finally, one last question concerns the estimate of the key quantum parameter γ. If this estimate is generally close to zero, then the quantum model reduces to a standard Markov model. Busemeyer et al. (2012) performed a hierarchical Bayesian estimation of the model parameters using the individual data in Barkan and Busemeyer (2003). The posterior group level distribution of gamma γ fell almost entirely below zero with a posterior mean equal to -2.67, which is clearly within the quantum regime.

Conclusions

This is the first time a quantum decision model was quantitatively compared to arguably the foremost standard and well-known decision model using a Bayesian model comparison method. A Bayes factor was computed, separately for each participant using both a uniform and a normal prior over the parameters. The Bayes factor, which takes both prediction accuracy and model complexity into account, clearly favored the quantum model, indicating a robust ability of the quantum model to make accurate predictions over a wide range of the parameter space.

The findings are specific to the examined experiment (one providing ample data that challenge many models) and specific to the models that we compared and their priors (the standard model is the best one known for fitting this particular data set, and the quantum model was matched to it in many respects). Nonetheless, the findings provide a first careful test and demonstration that the success of quantum models in accounting for human judgment and decision behaviors does not necessarily rely on greater model complexity. The generality of this conclusion of course will have to be explored in many other applications. The Bayes factor (and more generally, all measures of model complexity) depend not only on the model, but also on the experimental design and results (Myung et al., 2007). We do not intend to conclude from the present comparisons that quantum decision models are generally superior to standard decision models in all or most contexts. This is a question calling for continuing research to generalize the model comparison findings to other research paradigms, data sets, and models. We believe stronger evidence supporting the quantum model comes from applying the same general principles to a wide variety of research paradigms. Insofar, the exact same quantum model described here for the two-stage gambling task has been used to account for puzzling findings from completely different decision paradigms, including violations of the sure thing principle found with the prisoner's dilemma game (Pothos & Busemeyer, 2009) and interference effects found with a categorization–decision-making task (Busemeyer et al., 2009).

References

Aerts, D. (2009). Quantum structure in cognition. *Journal of Mathematical Psychology, 53,* 314–348.

Atmanspacher, H., & Roemer, H. (2012). Order effects in sequential measurements of non-commuting psychological observables. *Journal of Mathematical Psychology, 56,* 274–280.

Barkan, R., & Busemeyer, J. R. (1999). Changing plans: Dynamic inconsistency and the effect of experience on the reference point. *Psychological Bulletin and Review, 10,* 353–359.

Barkan, R., & Busemeyer, J. R. (2003). Modeling dynamic inconsistency with a changing reference point. *Journal of Behavioral Decision Making, 16,* 235–255.

Behme, C. (2013). Uncertainty about the value of quantum probability for cognitive modeling. *Behavioral Brain Science, 36,* 279–280.

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

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

Busemeyer, J. R., Wang, Z., & Trueblood, J. S. (2012). Hierarchical Bayesian estimation of quantum decision model parameters. In J. R. Busemeyer, F. DuBois, A. Lambert-Mogiliansky, & M. Melucci (Eds.), *Quantum interaction. Lecture notes in computer science* (Vol. 7620, pp. 80–89). Berlin, Germany: Springer.

Busemeyer, J. R., Weg, E., Barkan, R., Li, X., & Ma, Z. (2000). Dynamic and consequential consistency of choices between paths of decision trees. *Journal of Experimental Psychology: General, 129,* 530–545.

Cubitt, R. P., Starmer, C., & Sugden, R. (1998). Dynamic choice and the common ratio effect: An experimental investigation. *Economic Journal, 108,* 1362–1380.

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

Hey, J. D., & Knoll, J. A. (2007). How far ahead do people plan? *Economic Letters, 96,* 8–13.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. New York, NY: Cambridge University Press.

Jeffreys, W. H., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American Scientist, 80,* 64–72.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Society, 90,* 773–795.

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

Lambert-Mogiliansky, A., Zamir, S., & Zwirn, H. (2009). Type indeterminacy: A model of the "KT" (Kahneman-Tversky) man. *Journal of Mathematical Psychology, 53,* 349–361.

La Mura, P. (2009). Projective expected utility. *Journal of Mathematical Psychology, 53,* 408–414.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press.

Machina, M. (1989). Dynamic inconsistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature, 27,* 1622–1668.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. F. Manski & D. McFadden (Eds.), *Structural analysis of discrete data with economic applications* (pp. 198–272). Cambridge, MA: MIT press.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin and Review, 4,* 79–95.

Myung, I. J., Pitt, M. A., & Navarro, D. J. (2007). Does response scaling cause the generalized context model to mimic a prototype model? *Psychonomic Bulletin and Review, 14,* 1043–1050.

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,* 2171–2178.

Pothos, E. M., & Busemeyer, J. R. (2012). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences, 36,* 255–274.

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

Rakow, T. (2013). If quantum probability = classic probability + bounded cognition: Is this good, bad, or unecessary? *Behavioral Brain Science, 36,* 204–305.

Sarin, R., & Wakker, P. P. (1998). Dynamic choice and non-expected utility. *Journal of Risk and Uncertainty, 17,* 87–119.

Savage, L. J. (1954). *The foundations of statistics*. New York, NY: Wiley & Sons.

Shanteau, J., & Weiss, D. J. (2013). Physics envy: Trying to fit a square peg into a round hole. *Behavioral and Brain Sciences, 36,* 306–307.

Shiffrin, R. M., Lee, M., Kim, W., & Wagenmakers, E. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science, 32,* 1248–1284.

Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability model of causal reasoning. *Frontiers in Cognitive Science, 3,* 138.

Tversky, A., & Kahneman, D. (1990). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297–323.

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

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., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science, 5,* 672–688.

Yukalov, V. I., & Sornette, D. (2009). Physics of risk and uncertainty in quantum decision making. *European Physical Journal B, 71,* 533–548.

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

Received November 26, 2013
Revision received May 5, 2014
Accepted June 25, 2014 ■