

Supplemental Information for Bayesian comparison favors quantum over standard decision theory account of dynamic inconsistency

Jerome R Busemeyer
Indiana University

Zheng Wang
Ohio State University

Richard M. Shiffrin
Indiana University

April 5, 2014

Abstract

This supplement has five parts. The first part discusses the reasons for using a Bayes factor. The second part presents the 34 means choice probabilities from the Barkan and Busemeyer (2003) experiment. The third part summarizes the fits to the means of the models based on prospect theory and quantum theory. The fourth part describes the prior distributions used for the Bayesian model comparison. The fifth part summarizes the reasons for the behavior of the log likelihood function produced by the quantum model.

Keywords: Quantum probability, prospect theory, dynamic consistency, interference effects, Bayes factor.

1 Bayes Factor

That counting parameters is an inappropriate way to control for complexity is easy to illustrate with simple examples. Suppose a model has two parameters a and b , and AIC or BIC count these as two. But in fact the parameters obey the relation $c = ab$, so that a and b are perfectly (anti)correlated. In this case the ‘true’ parameter count would be one, not two. But putting aside correlations among parameters, parameters should not count equally. Imagine a data set with a thousand conditions, each data point representing the probability of a given response in one of these conditions. A model is applied and among its parameters are d and e . The value of d helps determine the predicted probability over a range of (0,1) in every condition. The value of e helps determine the predicted probability in only a single condition, over a range of (.6,.7). It should be obvious that these parameters should not be counted equally when assessing complexity. The converse of these examples also applies: one parameter can

represent or do the work of many. In the first example above, c does the work of a and b . It is also the case that one parameter can do the work of several uncorrelated parameters. In the second example above suppose the parameter e controls which of successive .10 regions contains the predicted probability, and ten parameters $e(n)$ control predicted probability within each of those regions. It is obvious that a single parameter d could cover the same (0,1) range of predictions. Another example: Suppose a model has two parameters a and b , each in the range (0,1) and each represented in practice by two decimal precision: $a = xy$ and $b = wz$. Let an equivalent model have a new parameter replacing a and b : $c = xywz$. Another equivalent model have a different new parameter replacing a and b : $c = 2x3y5w7z$. When c is ‘unpacked’ correctly either of the new models is equivalent to the first. These are all artificial examples, but serve to demonstrate that parameter counts are not the way to quantify complexity. In actual modeling these sorts of issues occur in many ways that are not at all trivial and artificial, and the way that parameters map into predictions for data can be very difficult to intuit. A true account of complexity requires an analysis of the way that the parameter space is mapped onto data outcomes. If for example all combinations of parameter values predict (roughly) the same data points, that model has very low complexity; if different parameter values produce different predictions, and if those predictions cover most of the possible outcomes that could have been observed, then a model has very high complexity. Bayesian Model Selection is the best current method for quantifying complexity and its balance with good fit (Minimum Description Length is a competitor for “best current method,” but in almost all cases gives answers very similar to those produced by Bayesian Model Selection).

Another question that can be raised is whether or not a Bayesian analysis can be performed using predictions derived from a quantum probability model. The answer to this question is yes for the following simple reason. For each condition, a sequence of outcomes are observed. The quantum model is used to assign a probability to each outcome sequence. The traditional decision model also assigns a probability to each outcome sequence. They assign different probabilities to these sequences using either quantum or classic probability rules. However, for each condition and for each model, there is a set of mutually exclusive and exhaustive sequence of outcomes. Both models assign a probability to each outcome sequence, which at this point obey the axioms of traditional Kolmogorov probability (e.g. the probability assigned to mutually exclusive outcome sequences are additive and the probabilities assigned to all possible outcome sequences sum to unity).

2 Barkan and Busemeyer (2003)

A two stage gambling paradigm was used to study dynamic consistency, which was a modification of the paradigm used by Tversky and Shafir (1992) to study the disjunction effect. A total of 100 people participated and each person played the 17 gambles involving real money shown in Table SI1 twice. Each gamble had

an equal chance of producing a win or a loss. The columns labeled “win” and “loss” indicate the money that could be won or lost for each gamble (one unit was worth one cent) and the column labeled EV shows the expected value of each gamble. For each gamble in Table SI1, the person was forced to play the first round, and then contingent on the outcome of the first round, they were given a choice whether or not to play the second round with the same gamble. On each trial the person was first asked to make a plan for the second play contingent on each possible outcome of the first play. In other words, during the planning stage they were asked two questions: “if you win the first play, do you plan to play the second gamble?” and “if you lose the first play, do you plan to play the second gamble?” Following the plan, the outcome of the first gamble was revealed, and then the person was given a final choice: decide again whether or not to play the second gamble after observing the first play outcome. To incentivize both plan and final choices, the computer randomly selected either the planned choice or the final choice to determine the real monetary payoff for each trial. The final payment for the trial was then shown to the person at the end of each trial. Participants were paid by randomly selecting four problems from the entire set, randomly selecting either their plan or final choice, and randomly selecting an outcome for each gamble to determine the actual payment.

Table SI1 displays the results obtained after averaging across the two replications for each person, and after averaging across all 100 participants. The columns under the label “Gamble” display the amount to win and lose for each gamble.

Table SI1: Barkan and Busemeyer (2003) Experiment							
Gamble		Win First Play		Gamble		Lose First Play	
Win	Loss	Plan	Final	Win	Loss	Plan	Final
200	220	0.46	0.34	80	100	0.36	0.44
180	200	0.45	0.35	100	120	0.47	0.63
200	200	0.59	0.51	100	100	0.63	0.64
120	100	0.70	0.62	200	180	0.57	0.69
140	100	0.62	0.54	160	140	0.68	0.69
200	140	0.63	0.53	200	160	0.67	0.72
200	120	0.74	0.68	160	100	0.65	0.73
200	100	0.79	0.70	180	100	0.68	0.80
				200	100	.85	.82

The probability of planning to take the gamble is shown under the column labeled “Plan.” There was little or no difference between the probabilities of taking the gamble, contingent each planned outcome of the first gamble, and so the results shown here are averaged across the two hypothetical outcomes during the plan. See Barkan and Busemeyer (2003) for the complete results listed separately for each contingent outcome. The probability of taking the gamble during the final stage is shown under the column labeled “Final.” Changes in probabilities down the rows of the Table show the effect of the gamble payoffs on the probability of taking the gamble. The difference between

the planned and final columns indicates a dynamic inconsistency effect. Notice that following a win (the first 4 columns), the probability of taking the gamble at the final stage was always smaller than the probability of taking the gamble at the planning stage. In other words, participants changed their minds and became more risk averse after experiencing a win as compared to planning for a win. Notice that following a loss (the last 4 columns), the probability of taking the gamble at the final stage was always smaller than the probability of taking the gamble at the planning stage. In other words, participants changed their minds and became more risk seeking after experiencing a loss as compared to planning for a loss.

3 Unitary rotation matrix used in quantum model

Evaluation of the payoffs causes the state ψ to be “rotated” by a unitary operator U into a decision state $\psi_D = U \cdot \psi_I$ used to make a choice about taking or rejecting the second stage gamble. The unitary operator is defined by

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

$$H_1 = \begin{bmatrix} \frac{h_W}{\sqrt{1+h_W^2}} & \frac{1}{\sqrt{1+h_W^2}} & 0 & 0 \\ \frac{1}{\sqrt{1+h_W^2}} & \frac{-h_W}{\sqrt{1+h_W^2}} & 0 & 0 \\ 0 & 0 & \frac{h_L}{\sqrt{1+h_L^2}} & \frac{1}{\sqrt{1+h_L^2}} \\ 0 & 0 & \frac{1}{\sqrt{1+h_L^2}} & \frac{-h_L}{\sqrt{1+h_L^2}} \end{bmatrix}, H_2 = \frac{-\gamma}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

The upper left corner of H_1 is defined by the payoffs given a win; and the bottom right corner of H_1 is defined by the payoffs given a loss. The parameters h_i , $i = W, L$, in H_1 are determined by the utilities of playing the gamble following the outcome of the first stage: $h_i = \tanh(.5 \cdot D_i)$, where $D_W = u(G|Win) - x_W^a$ is defined by Equations 2,3 in the article. For example, if $\gamma = 0$ and $h_W = h_L = h$, then the probability of taking the gamble is an increasing S-shaped function of h (see Pothos and Busemeyer, 2009; Busemeyer & Bruza, 2012, Ch. 9).

The matrix H_2 aligns beliefs with actions of the decision-maker by amplifying the potentials for states WT, LR and attenuating potentials for states WR, LT to produce what is called an entanglement state (see Pothos and Busemeyer, 2009; Busemeyer & Bruza, 2012, Ch. 9). In this application, entanglement refers to a state that is non-decomposable because of interdependence between beliefs about the first stage gamble and the intentions for actions to take during the second stage gamble. In particular, in this application, it captures the idea of a belief in a “hot hand,” e.g., winning the first hand is correlated with playing again on the second hand. The parameter γ adjusts the degree of changes in beliefs during the decision process. For example, if $H_1 = 0$, then $\gamma = 1$ produces

$|\psi_{WT}|^2 = .5 = |\psi_{LR}|^2$ and $|\psi_{WR}|^2 = 0 = |\psi_{LT}|^2$, producing an entangled state containing only WT, LR .¹

4 Model comparisons based on least squares fits to mean data

This section compares various versions of the two competing models based on least squares fits to the 34 means reported in Barkan & Busemeyer (2003, see SI). The models are compared using $R^2 = 1 - \frac{SSE}{TSS}$, and *adjusted* $R^2 = 1 - \frac{SSE}{TSS} \cdot \frac{34-1}{34-n}$, with $SSE = \sum (P_i - p_i)^2$, $TSS = \sum (P_i - \bar{P})^2$, P_i is the observed mean proportion of trials to choose gamble $i = 1, \dots, 34$, p_i is the predicted mean proportion, \bar{P} is the grand mean, $34 = 17$ (payoff conditions) $\times 2$ (plan vs final stage choices) is the number of means being fit, and n = number of model parameters.

4.1 Model R

Model R has $n = 3$ parameters, and the best fitting parameters (minimizing sum of squared error) are $a = .87$, $b = .92$, and $\gamma = 2.70$. The loss aversion parameter b is less than one (i.e., less sensitivity to losses), even though it is theoretically expected to be greater than one (i.e., greater sensitivity to losses). The model produced $R^2 = .77$ and an *adjusted* $R^2 = .76$. We also compared the *conventional* utility model defined in the article with an *alternative* utility model defined by $u(x) = x^a$ for $x \geq 0$ and $u(x) = |x|^b$ for $x < 0$. The alternative model also has $n = 3$ parameters, and the best fitting parameters are $a = .92$, $b = .79$, $\gamma = 2.47$. This alternative version of the reference point model also produced $R^2 = .77$ and *adjusted* $R^2 = .76$. Finally, we fit a reference point model using the conventional utility function that allowed for unequal weights for gains and losses. In this case, we replaced the .50 probability of a win with a decision weight parameter w , and the decision weight for the probability of a loss equaled $(1-w)$. The best fitting $n = 4$ parameters are $a = .85$, $b = 1.00$, $\gamma = 2.47$ and $w = .52$, and this model produced a slightly higher $R^2 = .80$ and an *adjusted* $R^2 = .78$.

4.2 Model Q

This quantum model has $n = 3$ parameters, and the best fitting parameters (minimizing sum of squared error) are $a = .71$, $b = 2.54$, and $\gamma = -4.40$. The risk aversion parameter is a bit below one as expected, and the loss parameter b exceeds one, as it should be. The model produced an $R^2 = .82$ and an

¹The model predictions for γ oscillate according to a sinusoidal function, so that if $H_1 = 0$, then $+\gamma$ gives exactly the same result as $-\gamma$.

adjusted $R^2 = .81$.² We also compared the *conventional* utility model defined above with the *alternative* utility model defined by $u(x) = x^a$ for $x \geq 0$ and $u(x) = |x|^b$ for $x < 0$. The alternative model also has $n = 3$ parameters, and the best fitting parameters are $a = .93, b = .25, \gamma = -4.40$. The alternative model produced a slightly higher $R^2 = .85$ and *adjusted* $R^2 = .84$. We also fit a quantum model using the conventional utility function that allowed the decision weight to change for gains and losses. In this case, we replaced the .50 probability of a win with a parameter w , and the probability of a loss equaled $(1-w)$. The best fitting model using $n = 4$ parameters produced estimates $a = .82, b = 2.53, \gamma = -4.37, w = .51$, and the model fit produced an $R^2 = .83$ and an *adjusted* $R^2 = .82$.

It is worth noting that γ is an important parameter in the quantum decision model. As explained above, it captures the theoretical idea that people correlate their beliefs about winning the first stage gamble with playing again on the second stage. If we force $\gamma = 0$, then the quantum model is reduced to a Markov model, and we can no longer account for the violation of the “sure thing” principle described earlier (Pothos & Busemeyer, 2009). In the current study, $\gamma = 0$ means the choice probability for the plan is an equal weight average of the two choice probabilities produced after either winning or losing the first stage: $p(T|plan) = (.50) \cdot p(T|Win) + (.50) \cdot p(T|loss)$, where $p(T|j)$ is defined by Equation 4 with $\gamma = 0$. This model was fit to the results in Table SI1 by using only $n = 2$ parameters, a and b , for the quantum model (with $\gamma = 0$), and it produced an $R^2 = .78$ and an *adjusted* $R^2 = .78$.

Comparing the various versions of Model R with the corresponding versions of Model Q with respect to least squares fits to the means, we see that Model Q always performs better than the corresponding Model R by a few percentage of variance explained even though both models use the same number of parameters. However, best fits to aggregate data does not imply best fit to individual data, and so the next section examines fits at the individual level.

5 Model comparisons based on log likelihood fits to individual data

Next, we compare the two competing models based on log likelihood fits to data from each individual using the four combinations of plan and final choices observed from 33 gambles that each person completed. The definition of the log likelihood for each person is described in the article. The log likelihood from the each person was converted into $G^2 = -2 \cdot \ln(L_i)$ which indexes *lack* of fit, and the parameters that minimized G^2 were found for each person.³

²One result of the oscillating nature of the quantum model is that there are many local maximum. For example, the parameters $a = .60, b = 4.60, \gamma = 2.63$ produce an equally good fit with $R^2 = .82$.

³A surprising feature was found with the log likelihood function of the quantum model as a function of the key quantum parameter γ . The log likelihood function forms a damped oscillation that converges at a reasonably high log likelihood at the extremes, and this is

5.1 Model R

First we examined Model R using the *conventional* utility function. This model has $n = 4$ parameters for each person $\{a, b, \gamma, m\}$. The mean estimates, averaged across 100 individual estimates, were $a = .7067, b = .9095, m = .5789, \gamma = 3.09$. Summed across the 100 participants, this model produced a total $G^2 = 7020$. We also examined Model R using the *alternative* utility function defined by $u(x) = x^a$ for $x \geq 0$ and $u(x) = |x|^b$ for $x < 0$, which also has 4 parameters, but this model produced a larger total lack of fit with 4.

Next we examined a version of Model R that introduced an additional free parameter for the weight given to gains w and a weight $(1 - w)$ for losses. This model produced a total $G^2 = 6936$. This unequal decision weight model has $n = 5$ parameters, and so we used a difference in Bayesian Information Criterion (pooled across all 100 participants) to compare the two models $BIC\ diff = (BIC_{5\ Par} - BIC_{4\ Par})$. This model comparison produced $BIC\ diff = 725.60$, favoring the simpler $n = 4$ parameter model. If instead we perform a significance test, then the G^2 difference = 85 is less than the critical value for a .05 level of significance ($\chi^2(100) = 124$).

5.2 Model Q

Model Q was fit using the conventional utility function and this model also has $n = 4$ parameters $\{a, b, \gamma, m\}$. The mean estimates from fitting the 100 individuals were $a = .6518, b = 1.97, m = .5932, \gamma = -2.67$. Summed across the 100 participants, this model produced a total $G^2 = 6883$, which is an improvement over the four parameter version of Model R using the conventional utility function. A total of 77 out of the 100 participants produced better fits (as measured by G^2) for the four parameter version of Model Q as compared to the four parameter version of Model R using the conventional utility function.

In summary, the log likelihood fits and BIC comparisons indicate that the 4 parameter version of model R using the conventional utility function is the best version of model R. This version of the reference point model performs worse with respect to total G^2 lack of fit (using the same number of parameters) than the quantum model.

6 Prior distributions and Bayes factors

Two different prior distributions were examined: a uniform and a normal distribution. The uniform distribution assigned equal probability to each grid combination point (described in the article). For the normal prior, we assumed independent (discretized) normal distributions for each parameter denoted as

true both for the average across participants as well as for individual participants. See SI for details.

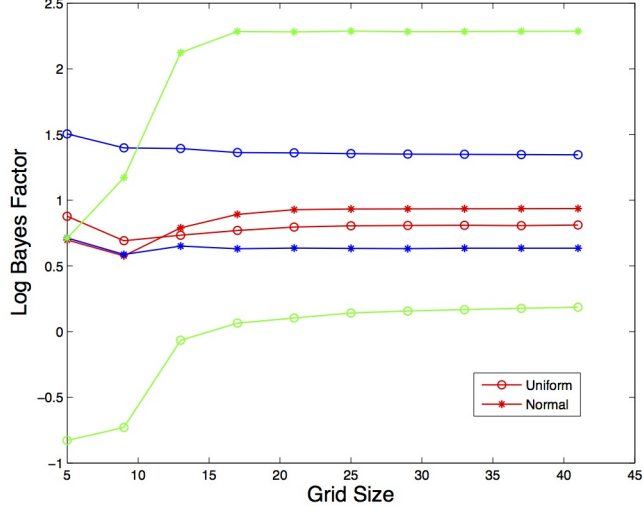


Figure 1: Convergence of Bayes Factor

$f(x_i)$ for the i -th value of one of the parameters.

$$g(x_i) = e^{-\left(\frac{x_i - \mu}{\sigma}\right)^2}, f(x_i) = \frac{g(x_i)}{\sum g(x_i)}, i = 1, 41.$$

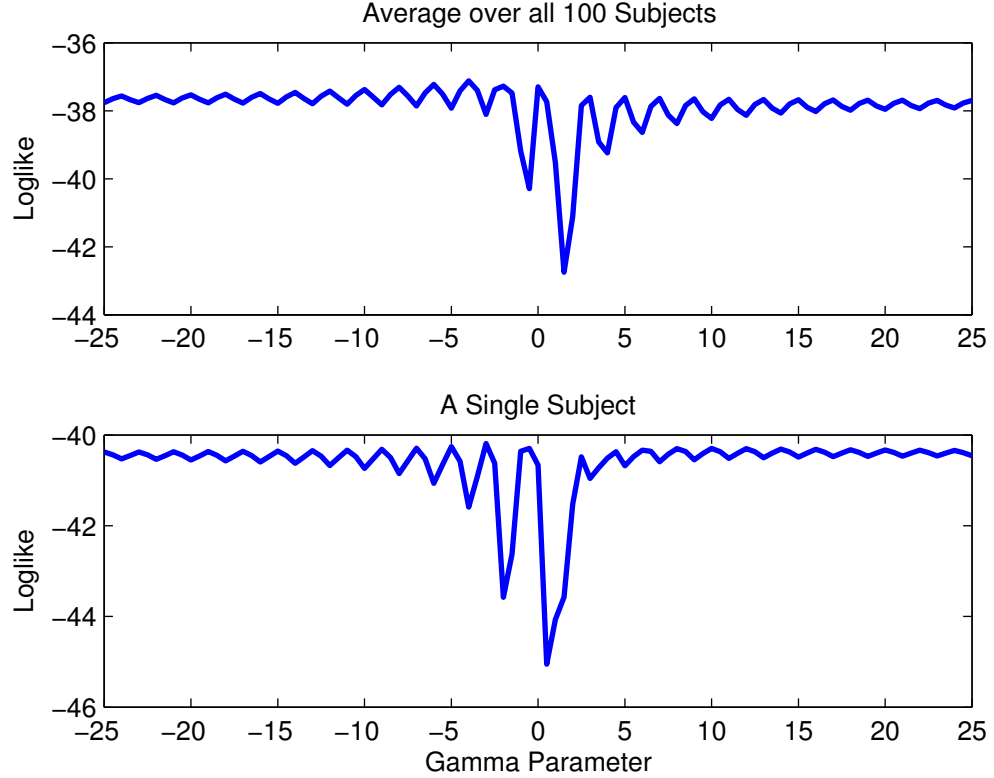
The means and standard deviations used for the comparisons are described in the article. We examined convergence of the Bayes factor as a function of the number of grid points. This is shown in Figure 1 for three different participants.

In addition to the Bayes factor comparisons described in detail in the article, we also conducted several other comparisons between model R versus model Q using the conventional utility function. In one analysis, when using the uniform prior, we reduced the ranges of the parameters to one half the range described in the manuscript, and in this case the total log Bayes factor equaled 98.85 favoring the quantum model. In another analysis, we reduced the standard deviations as follows: standard deviation for a equal to .25, standard deviation for b equal to .25, standard deviation of m equal to .25, standard deviation of γ for Model R equal to .75, and standard deviation of γ for model Q equal to 5. The latter produced a total log Bayes factor equal to 159.93 in favor of model Q.

7 Log likelihood as a function of gamma

A surprising feature was found with the log likelihood function of the quantum model. Figure 2 plots the log likelihood (averaged across participants and other

Figure 2: Log likelihood as a function of Gamma



parameters) for the quantum model when the parameter γ varies across a wide range. As can be seen in the figure, the log likelihood function forms a damped oscillation that converges at a reasonably high log likelihood at the extremes, and this is true both for the average across participants as well as for individual participants.

Why does the log likelihood function for the quantum model reach a reasonably high asymptote as the quantum parameter gamma increases in magnitude in either the positive or negative direction?

The log likelihood converges to an asymptote because the model predictions always converge to an asymptote for all model parameter values. Why do the model predictions converge to an asymptotic value as gamma increases?

The quantum model predictions are derived from a Hamiltonian (which is real and symmetric) that can be decomposed into four real eigenvalues and four real valued orthonormal eigenvectors. So the predictions are computed from these eigenvalues and eigenvectors. The eigenvalues come in pairs – plus-minus for one value, and plus-minus for a second second value.

As gamma increases, the 4 eigenvalues converge to 4 values that vary with gamma: $[(-\text{gamma} - \text{constant}), (-\text{gamma} - \text{constant} + \text{difference}), (\text{gamma} + \text{constant} - \text{difference}), (\text{gamma} + \text{constant})]$. The term “constant” is a number that is fixed and does not depend on gamma. The term “difference” is the difference between the first and second eigenvalue, and the difference between the third and fourth eigenvalue. This difference also converges to a fixed value as gamma increases. In the formula used to compute the probabilities for each action, we can factor out gamma so that it disappears in the computation of the probabilities – except for one term, which is a cross product interference type of term. Gamma has a multiplicative effect on this cross product term. The cross product term is composed of sums of products of the values in the eigenvectors. However as gamma increases, this cross product interference term approaches zero. Why?

As mentioned earlier, the eigenvalues come in pairs, plus - minus for one value; plus - minus for another. The eigenvectors are determined from the eigenvalues, and consequently, the two eigenvectors for a pair have four values that are the same magnitudes, but they appear in permuted positions and only change in sign.

As gamma increases, the eigenvalues all increase, but the value of the eigenvalue used for one pair becomes approximately the same as the value of the eigenvalue used for another pair. Once again, the eigenvalues are used to determine the eigenvectors. As gamma increases, all four eigenvectors converge to have the same magnitudes, and they only differ in terms of the permutation of the values and the signs. In other words, all four eigenvectors can be formed by permuting the first one and changing signs of the first one. This produces a very constrained set of eigenvectors. Recall that the cross product term is formed by the sum of products of the values of the eigenvectors. The constrained values of the eigenvectors satisfy precisely the property required to force the cross product interference term to zero. The only effect of gamma is produced by multiplying the cross product term, and so when this term goes to zero, the effect of gamma disappears.

References

- Barkan, R., & Busemeyer, J. R. (2003). Modeling dynamic inconsistency with a changing reference point. *Journal of Behavioral Decision Making*, 16, 235-255.
- Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science*, 3, 305-309.